FEREDIT
Fondo editorial
Red de investigadores de la Transcomplejidad

Estadística transcompleja

Inferencia, ética y algoritmos en la nueva ciencia

Leslibeth Joseliz Sucre González, Carluys Suescum Coelho, Car-Emyr Suescum Coelho, María Gabriela Gutiérrez Cadenas, Thania Jheny Torres Pernía, Yasmilys Graciela Briceño Pérez

Estadística transcompleja: inferencia, ética y algoritmos en la nueva ciencia

Leslibeth Joseliz Sucre González, Carluys Suescum Coelho, Car-Emyr Suescum Coelho, María Gabriela Gutiérrez Cadenas, Thania Jheny Torres Pernía, Yasmilys Graciela Briceño Pérez.

Colección: **Praxeología y Transmetodología de la Transcomplejidad**
Primera Edición, Enero, 2026
Depósito Legal: **AR2026000012**
ISBN: **978-980-456-061-3**

Reservados todos los derechos conforme a la ley
Se permite la reproducción total o parcial del libro, siempre que se indique expresamente la fuente

Libros@Red de Investigadores de la Transcomplejidad.
https://reditve.wordpress.com
Rif: J403566976
Portada: Leslibeth Sucre y Thania Reyes

AUTORIDADES REDIT

Dra. Crisálida Villegas
Presidente
Dra. Nancy Schavino
Vicepresidente
Dra. Mary Stella
Directora de Administración
Dra. Alicia Uzcátegui
Secretaria

FEREDIT
Dra. Sandra Salazar
Directora

Comité Editorial
Dra. Betty Ruiz
Dra. Rosana Silva
Dra. Evelyn Ereú
Dra. Miozotis Silva
Dr. Renne Pérez
Dr. Arturo Dávila

ÍNDICE DE CONTENIDO

		pp.
	Presentación Leslibeth Joseliz Sucre González	7
I	La cartografía ética del criterio humano en la inferencia algorítmica Leslibeth Joseliz Sucre González	12
II	Más allá del p-valor: inferencia práctica y decisiones Carluys Suescum Coelho	26
III	De la asociación a la causalidad en investigación cuantitativa Car-Emyr Suescum Coelho	42
IV	Recolección de datos para pruebas de hipótesis en investigación María Gabriela Gutiérrez Cadenas	66
V	Algoritmos aumentados: la nueva frontera de la inferencia cuantitativa tabular Thania Jheny Torres Pernía	92
VI	El número de Harshad-Niven y su aplicabilidad estadística en la investigación Yasmilys Graciela Briceño Pérez	110
	Referencias	127

ÍNDICE DE TABLAS

No.		pp.
1	Medidas de asociación frecuentes	57
2	Riesgos de error y controles de recolección en pruebas de hipótesis	87
3	Algoritmo aumentado como un sistema	107

ÍNDICE DE FIGURAS

No.		pp.
1	Dualidad de la inferencia: de la base clásica a la predicción digital.	17
2	Mapeado desde la dimensión valorativa del Ser	21
3	Tipo de investigación según el propósito	45
4	Diagrama de Poincaré para la secuencia MRH	123
5	Divergencia logarítmica de trayectorias vecinas	125

PRESENTACIÓN

La investigación en esta nueva época científica transformó su transitar por un camino lineal de verdades inalterables hacia un trayecto de navegación consciente y reflexiva que reconoce y abraza la incertidumbre; además, distingue que la realidad no es un objeto inerte que espera ser cuantificado, sino que la visualiza como una estructura viva, dinámica y en red.

Desde esta perspectiva, es imposible mirar la estadística desde el ritual mecanicista del siglo pasado; esta precisa una resignificación para elevarse a la hermenéutica de datos capaz de emprender un diálogo con la ética, la tecnología y las reflexiones humanas.

El libro **Estadística Transcompleja: Inferencia, Ética y Algoritmos en la Nueva Ciencia** surge como una respuesta activa y epistémica frente a los retos de gestar procesos de investigación en la era del Big Data y la Inteligencia Artificial.

Este libro que es un trabajo sinérgico y colaborativo de seis investigadores no es un manual de procedimientos estadísticos; su propósito es una invitación a repensar los procesos lógicos empleados en la producción y generación de conocimiento.

Las páginas de la obra plasman las ideas desafiantes de los autores, desarmando la falsa seguridad de la certeza absoluta en la hegemónica forma de hacer ciencia para dar lugar a una nueva visión infinitamente más sólida y humana, caracterizada por elevar la humildad epistémica y la justicia cognitiva.

La obra se estructura en seis capítulos que transitan una ruta renovadora cuyo exordio inicia con la fundamentación filosófica, epistémica y ética, continúa con miradas críticas y prácticas de la inferencia estadística hasta reconocer y asumir la actual frontera tecnológica y sus aplicaciones especializadas.

En el primero, *Cartografía ética del criterio humano en la inferencia algorítmica*, la autora plantea un aporte esencial, ante la opacidad del algoritmo y la indiscutible velocidad de las herramientas tecnológicas, pues resignifica la phrónesis y el juicio axiológico del investigador como el único medio para perpetuar el sentido y la humanidad en la ciencia.

El segundo, por su parte, *Más allá del p-valor: inferencia práctica y decisiones,* desafía la hegemonía del ritual estadístico, ya que el autor plantea integrar tamaños del efecto e intervalos de confianza para ir más allá de la mejora de la precisión técnica y contribuir a la necesidad de relevancia y pertinencia social en el marco del ODS 17.

El tercero, "*De la asociación a la causalidad en investigación cuantitativa",* expone una idea renovadora para un nodo crítico de la investigación cuantitativa; el autor presenta criterios preventivos

sobre la extrapolación no justificada desde una argumentación causal clara.

Seguidamente, en el cuarto, *Recolección de datos para pruebas de hipótesis*, la autora plasma una reflexión profunda al referir que los procesos de análisis sofisticados se consideran totalmente improductivos y estériles si no existe garantía de la tríada operativa esencial en la investigación: calidad, validez y confiabilidad del dato.

Por su parte, en el quinto, *Algoritmos aumentados: la nueva frontera de la inferencia cuantitativa tabular,* la autora explora los algoritmos aumentados y muestra una nueva frontera de la inferencia tabular, así como de la "agentización" del ciclo estadístico, en la que la Inteligencia Artificial se reconoce como organizadora asertiva de procesos.

Finalmente, en el sexto, *El número de Harshad-Niven y su aplicabilidad estadística en la investigación*, las ideas de la autora demuestran cómo las distribuciones matemáticas no lineales generan una ruptura en la clásica predictibilidad que

se transmuta en entropía y seguridad en los procesos de investigación.

Esta obra promulga que la estadística transcompleja se considera un ejercicio de responsabilidad intelectual, pues su acción no es simplemente establecer datos a través de cálculos de valores y números para simplemente predecir; su acción es comprender e inferir con conciencia ética.

El libro representa una brújula para el investigador contemporáneo cuyo propósito es trascender la sencillez de un dato frío y reconocer la profundidad de una nueva ciencia permeada de ética, equidad y sentido humano.

Leslibeth Sucre González

I. LA CARTOGRAFÍA ÉTICA DEL CRITERIO HUMANO EN LA INFERENCIA ALGORÍTMICA

Leslibeth Joseliz Sucre González[1]

La investigación científica actualmente se encuentra en un punto de inflexión metodológica, en el que la incertidumbre se impone a la certeza y el dato se entrelaza con la interpretación. En este sentido, es necesario reconocer que la estadística aplicada no es simplemente un conjunto de herramientas de cálculo, sino la lógica formal para cuantificar la incertidumbre.

La inferencia algorítmica es un proceso sofisticado que combina la lógica estadística y computacional para identificar patrones, hacer generalizaciones, plantear conclusiones o prever

[1] Doctora en Educación. Postdoctora en Investigación Universitaria. Académico responsable de postgrado de la Universidad Nacional Abierta. Profesora de la Universidad Pedagógica Experimental Libertador. leslibethsucregonzalez@gmail.com. 0009-0000-5867-7020

resultados a partir de grandes volúmenes de datos (Big Data).

Utiliza modelos matemáticos complejos que son ejecutados por algoritmos de inteligencia artificial (IA) y aprendizaje automático (machine learning); esto le ha permitido convertirse en una herramienta innovadora para generar conocimiento.

En relación con esto, López et al. (2025) plantean que la estadística trasciende su función descriptiva para actuar como mediadora en la generación de conocimiento, optimizando la toma de decisiones mediante evidencia comprobable y fomentando la mejora continua en diversos campos (p.6).

Ahora bien, delegar la inferencia, ese procedimiento que va del dato específico a la regla general, que implica la responsabilidad de realizar deducciones, interpretar información y extraer conclusiones en sistemas autónomos, plantea una profunda reflexión ética y ontológica, surgiendo una interrogante esencial desde la práctica de la

investigación contemporánea ¿Puede el algoritmo, sin axiología, sustituir la phrónesis humana, que da sentido a la cuantificación?

El propósito de este capítulo, es propiciar una deliberación adecuada para delinear un mapa del ámbito en el que la eficacia opaca de la máquina se une al juicio ético del investigador, defendiendo la huella ineludible que deja el ser humano al crear el conocimiento.

Horizonte epistemológico de la inferencia algorítmica y la probabilidad

Desde su esencia, la probabilidad representa el lenguaje de la incertidumbre; al respecto, Serrano-Díaz (2022) refiere que “la probabilidad, por ser de naturaleza matemática, está permeada de intuición autárquica, utilidad práctica y teorización” (p.5)

De forma tradicional, la inferencia estadística se reconoce desde dos amplios enfoques: el primero, el frecuentista u objetivista; este representa la base de la estadística clásica que destaca los intervalos de confianza y p-valores.

Además, se fundamenta en el límite de la frecuencia relativa. La función consiste en medir de forma objetiva el riesgo de cometer errores al rechazar una hipótesis. En relación con la inferencia estadística clásica, Toala et al. (2024) declaran que es el procedimiento por el cual, a partir de los valores de una muestra, se determinan los valores esperados en la población bajo ciertas circunstancias y se generan pronósticos sobre qué tan verdaderos pueden considerarse dichos valores.

El segundo enfoque se denomina bayesiano, también conocido como epistemológico/subjetivo; este se basa en el nivel de creencia, que se actualiza constantemente con nuevos datos a través del teorema de Bayes.

Ramos-Vera (2020) describe que la estadística bayesiana, permite la cuantificación de las pruebas vinculadas con las hipótesis y los modelos asociados, lo cual vigoriza la metodología estadística y proporciona estimaciones precisas y contundentes.

Adicionalmente, este enfoque emplea el factor Bayes como un medio cuantificable para determinar hasta qué punto los datos apoyan dos modelos de hipótesis. En este procedimiento, los valores interpretables se organizan en: "débil", "moderado", "fuerte" y "muy fuerte".

Los investigadores en los diferentes ámbitos de la sociedad actual emplean recursos desde la inteligencia artificial (IA) como apoyo del proceso investigativo. En este sentido, es propio referir que la IA funciona, principalmente, mediante modelos de optimización complejos o lógicas bayesianas amplificadas que, aunque son rigurosos desde el punto de vista matemático, tienden a separarse del análisis ontológico.

La inferencia algorítmica contemporánea consiste en la habilidad de deducir patrones y realizar predicciones a una velocidad y en una magnitud que superan la capacidad humana. No obstante, este ritmo puede generar el peligro de la opacidad algorítmica, en la que se pierde el "por qué"

de una conclusión en la estructura del modelo predictivo o de la red neuronal.

En la figura 1, a continuación, se agrupan los principales enunciados antes descritos en este horizonte epistemológico de la inferencia algorítmica y la probabilidad.

Figura 1. Dualidad de la inferencia: base clásica y predicción digital
Fuente: Sucre (2025) generada con Gemini

La figura 1 establece el debate principal de este apartado referido a cómo la investigación, al transitar de las bases firmes del enfoque frecuentista y las creencias actualizables desde el enfoque bayesiano

al modelo predictivo de la IA, requiere con urgencia la phrónesis humana para preservar su capacidad explicativa ante la carencia de transparencia del algoritmo y su coherencia ética.

Criterio humano: la cartografía axiológica

El criterio humano va más allá de la simpleza del dominio instrumental de realizar cálculos o de manejar software estadístico avanzado, es un acto con matices axiológicos, creativos y contextualizados.

Al respecto Barrio (2022) señala que el humano es un ser de razón y esta no solamente tiene valor para contar; sino que, es una habilidad para conocer y comprender desde captar la realidad de las cosas, dentro de las limitaciones inherentes, incluidas las propias restricciones del saber humano, sobre todo el hecho de que no tenemos un conocimiento pleno ni total.

El algoritmo carece de ética intrínseca y, por ende, de axiología. Es el investigador quien aporta los valores de la dignidad humana, la equidad, la

justicia y la pertinencia social. Entonces, es posible afirmar que es el quien determina si una variable es empleada desde lo ético o si un resultado es socialmente aceptable, esto, más allá de la simple significación estadística.

En este sentido, acotan Díaz y García (2024) que en la ciencia el rol del investigador es esencial, porque su influencia se extiende a la producción de conocimiento y, por ende, a la sociedad. El rol del investigador debe caracterizarse por una profunda responsabilidad ética.

Adicionalmente, los prenombrados autores señalan que la ética es esencial para guiar el comportamiento del investigador; por ello, es indispensable que este reconozca que su rol en el proceso de investigación no solo tiene una finalidad científica, sino también un impacto importante en la gente y las comunidades.

La inferencia algorítmica aporta al proceso de investigación correlaciones de datos y predicciones de acuerdo con los resultados, mientras que el

criterio humano aporta comprensión. Esta implica una reflexión profunda, histórica y cultural que únicamente el investigador puede aportar por estar sumergido en su objeto de estudio.

Un ejemplo palpable, en el ámbito educativo, es que una predicción de fracaso o deserción escolar requiere una intervención ética y social que, sin la comprensión que brinda el criterio humano, el algoritmo jamás podrá generar por sí solo.

La figura 2 representa un mapa desde esa dimensión valorativa que otorga el criterio humano.

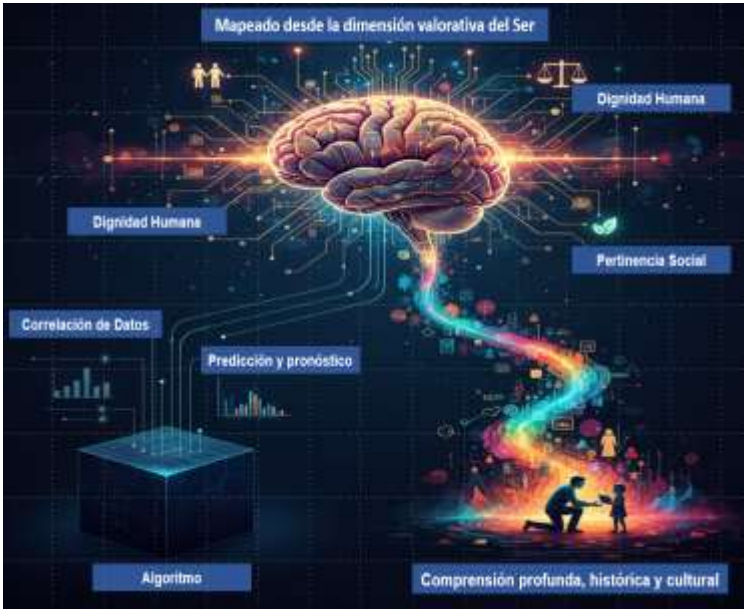

Figura 2. Mapeado desde la dimensión valorativa del ser
Fuente: Sucre (2025) generada con Gemini

La figura establece cómo el ser humano, desde la axiología que representa esa dimensión de valores éticos, logra mapear su realidad, en la cual permean sus experiencias, vivencias, decisiones, conductas y saberes, empoderándose de esos valores como una brújula para comprender el entorno que lo rodea, desde el cual se reinterpretan la historia y la cultura.

Cruce de caminos ético: inferencia, sesgo y opacidad algorítmica

En ese cruce de caminos de la tríada inferencia, sesgo y opacidad del algoritmo, es preciso reconocer, comprender y trazar los riesgos y las desviaciones que ocurren cuando el criterio y el rigor en la investigación se vuelven borrosos; es esencial entonces la intervención de la cartografía ética.

La inferencia algorítmica enfrenta como reto inminente la falta de imparcialidad inherente a los datos y a los modelos que los procesan. En este sentido, el sesgo característico se refiere a que los modelos de inteligencia artificial no son objetivos; dependen de la calidad de la información que se les suministra.

Zhai et al (2024) identificaron que los sesgos algorítmicos con frecuencia se generan porque los sistemas de inteligencia artificial se alimentan de datos con prejuicios inherentes, lo que genera que

los usuarios crean erróneamente que esos resultados sesgados son válidos, objetivos y reales.

Ante la posible perpetuación de los sesgos que el algoritmo puede reproducir y, más allá, que la inferencia algorítmica genere conclusiones que puedan lograr desde la adopción sin validación del sujeto, que se acentúen verdades inexistentes, es preciso que el investigador asuma el deber ético de limpiar, verificar y contextualizar los datos.

Es necesario reconocer que el manejo innovador de herramientas de inteligencia artificial permite gestionar grandes volúmenes de datos. Sin embargo, esto ha generado una profunda dependencia metodológica que pone en riesgo la existencia del pensamiento humano desde lo divergente y lo crítico.

Por ello, es inminente que el investigador se convierta en una especie de mediador ontológico para mantener una labor investigativa desde una perspectiva flexible y amplia, sin condicionarse a la información que genera la inteligencia artificial.

El pensamiento medular de esta reflexión capitular es reconocer que la cartografía ética del criterio humano exige una gobernanza algorítmica responsable en la investigación científica. Por ello, es necesario establecer el límite fronterizo entre la eficiencia de la inteligencia artificial y sus algoritmos y la esencia del criterio humano.

La inferencia algorítmica se reconoce como una herramienta poderosa para pronosticar en la incertidumbre; sin embargo, la ausencia de axiología en su accionar solo le permite actuar como un espejo probabilístico del contexto. Es por lo que la responsabilidad de aportar creatividad, ética, contexto y sentido humano es únicamente de ese criterio que puede aportar el investigador, más aún uno que asume la visión transcompleja del mundo.

Es esencial que la inteligencia artificial permanezca como un soporte, una guía y apoyo metodológico, pero que no se convierta en un simple sustituto del pensamiento crítico, divergente, flexible, creativo y genuino que caracteriza al ser humano.

Él es el auténtico artista generador de conocimiento; solo así la estadística aplicada a la educación dejará de ser una herramienta fría para transmutar y convertirse en un medio ético y esencial en la comprensión de la realidad humana.

II. MÁS ALLÁ DEL P-VALOR: INFERENCIA PRÁCTICA Y DECISIONES

Carluys Suescum Coelho[2]

A lo largo de muchas décadas, la significancia asintótica bilateral, mejor conocida como el p-valor, ha ocupado un lugar casi hegemónico y sempiterno en la investigación científica. En muchos campos, el "tener $p < 0,05$" se ha convertido en sinónimo no solo de descubrimiento, sino también, de mérito publicable.

Sin embargo, la evidencia acumulada ha demostrado que la dependencia casi exclusiva de este indicador ha conllevado a conclusiones que no solo pueden catalogarse de engañosas, sino también de decisiones poco informadas, no en vano, la American Statistical Association, mejor conocida como ASA por sus siglas en inglés, ha advertido que

[2] Doctor en Ciencias Gerenciales. Centro de Estudios Gerenciales Avanzados (CEGA). carluyscoelho@gmail.com . https://orcid.org/0009-0000-2044-7684

las decisiones científicas no deberían basarse únicamente en si un p-valor supera un umbral arbitrario, de acuerdo con Wasserstein & Lazar (2016).

Amrhein et al (2019) en sus investigaciones, han recogido las peticiones de un grupo conformado por más de 800 investigadores que exigen abandonar el uso de la significación estadística como criterio dicotómico. En paralelo, la Agenda 2030 para el Desarrollo Sostenible de las Naciones Unidas enfatiza la necesidad de generar sistemas estadísticos más sólidos y medidas de avance que vayan más allá de simples métricas orientadas a determinar el PIB.

Es aquí donde la meta 17.19 correspondiente al ODS 17 propone desarrollar mediciones del avance en el desarrollo sostenible que complementen el producto interno bruto y apoyar la creación de capacidad estadística en los países en desarrollo, de acuerdo con las Naciones Unidas (2025).

Partiendo de este panorama, al seguir atados a una lectura pobre de la evidencia que se circunscriba a una mera reducción de p-valores, resulta contradictorio con el mandato y necesidad de producir indicadores que sean cada vez más transparentes y pertinentes para la toma de decisiones.

Este capítulo tiene como objetivo mostrar que interpretar de forma integrada los tamaños de efecto y la incertidumbre (intervalos de confianza, bandas de credibilidad, escenarios probabilísticos), en lugar de limitarse a la mera significancia estadística, produce conclusiones mucho más útiles para la toma de decisiones reales.

La motivación de este giro metodológico está en consonancia con el espíritu de la meta 17.19 que exige la consolidación de capacidades estadísticas que sean capaces de producir indicadores robustos y sustantivamente relevantes para el seguimiento de los Objetivos de Desarrollo Sostenible (ODS).

Limitaciones de la hegemonía del p-valor

El p-valor indica cuán incompatibles son los datos observados con una hipótesis nula dada, este valor no mide ni la magnitud del efecto, ni mucho menos su importancia práctica, por ello, Mendivelso & Rodríguez (2018) señalan que la interpretación errónea que se realiza del p-valor y de los intervalos de confianza ha generado la falsa sensación de encontrar resultados con relevancia clínica siempre y cuando estos sean reportados como estadísticamente significativos.

Ahora bien, el problema se agrava cuando se convierte un umbral convencional ($\alpha = 0,05$) en una frontera sumamente rígida entre el “existe efecto” y “no existe efecto”, ignorando casi de manera completa no solo el contexto, sino el tamaño de la muestra y la variabilidad de los datos.

Esta lectura dicotómica tiene consecuencias, las cuales a lo largo del tiempo han sido bien documentadas por cuanto invisibiliza la magnitud de los fenómenos, incentiva malas prácticas como el p-

hacking (manipulación de resultados) y el sesgo de publicación.

Todo lo que en conjunto termina por conducir a descartar hallazgos potencialmente relevantes simplemente porque el p-valor no cruzó el umbral arbitrario. De hecho, en evaluaciones de políticas públicas, intervenciones educativas o programas de salud en territorios con una marca heterogeneidad, esta lógica dualista es sumamente dañina.

Para abordar y hacer frente a estas limitaciones, diversos autores han propuesto desplazar el foco desde el "es significativo" hacia el "cuán grande es el efecto, cuán incierto es y qué, efectivamente implica en la práctica". Es bajo esta lógica de las nuevas estadísticas donde Cumming (2014) sostiene que se debe priorizar la estimación y la comunicación de la incertidumbre sobre el ritual de las pruebas de significación.

Tamaño del efecto: de la significación estadística a la sustantiva

El tamaño del efecto constituye la medida cuantitativa de la magnitud de un fenómeno, por ejemplo, las diferencias de medias estandarizadas, las proporciones de varianza explicada o las razones de las probabilidades (odds ratios).

Esto porque a diferencia del p-valor, que solo informa sobre la compatibilidad de los datos con la hipótesis nula, el tamaño del efecto responde a una pregunta mucho más sustantiva: ¿cuán grande es el cambio, la diferencia o la asociación que observamos?

Coe & Merino-Soto (2003) explican que la medida de tamaño del efecto describe la magnitud de una diferencia y no simplemente si esa diferencia es estadísticamente significativa. Para ello, destacan especialmente que estas medidas permiten comparar resultados entre estudios y contextos.

En la práctica, destacan los autores, que la distinción entre un efecto pequeño (por ejemplo, d ≈ 0,20), mediano (d ≈ 0,50) o grande (d ≥ 0,80) ofrece a los tomadores de decisiones una pauta clave sobre

la relevancia potencial de una intervención, siempre que se parta de la premisa de que estos valores se interpreten a la luz del contexto disciplinar y de los umbrales de cambio considerados como socialmente significativos.

La investigación de Rodríguez-Sabiote et al (2025) va en la misma dirección, en el sentido de que mostraron mediante un ejemplo con el software JASP cómo la decisión basada solo en p-valores puede entrar en una franca contradicción con la interpretación basada en tamaños de efecto.

En su estudio destacan dos principales debilidades de la significación estadística: la dependencia del tamaño muestral y la arbitrariedad del umbral del valor p. Mientras que en contraposición el tamaño del efecto se presenta como una alternativa más robusta y explicativa.

Es así como una política pública, un programa educativo, una intervención en salud o una política social no debería juzgarse por si logra $p < 0,05$, sino

por la magnitud del verdadero cambio que genera en resultados realmente relevantes para las personas.

Incertidumbre e intervalos de confianza: haciendo visible la precisión

Un segundo aspecto clave de una inferencia orientada a la toma de decisiones es la explicitación de la incertidumbre, precisamente porque todo estimador, ya sea una media, una diferencia o un coeficiente de regresión, está sujeto al error muestral y a la variabilidad propia del fenómeno estudiado.

Es precisamente bajo este escenario que los intervalos de confianza permiten traducir esta idea en un rango de valores plausibles para el parámetro poblacional. Escrig Sos et al (2007) en su investigación, recuerdan que los intervalos de confianza no solo permiten primeramente valorar la significación estadística, sino también la significación clínica o práctica del resultado, porque muestran de una forma directa la amplitud del efecto compatible con los datos.

Los estudios más recientes han insistido en que un intervalo de confianza bien interpretado es mucho más informativo que un p-valor aislado, en este sentido, Mendivelso & Rodríguez (2018) señalan que interpretar un resultado no significativo directamente como una ausencia de efecto es un error, pues un intervalo amplio puede incluir tanto efectos beneficiosos relevantes como efectos nulos o incluso perjudiciales, es decir, lo que está indicando es simplemente que la evidencia disponible resulta insuficiente.

En los procesos que implican la toma de decisiones, esto es crucial, por ejemplo, si lo ilustramos con el caso de un programa de formación docente que muestre una mejora media de 0,30 desviaciones estándar con un intervalo de confianza (0,05; 0,55) ofrece un panorama distinto al de un programa con la misma media, pero con un intervalo (–0,10; 0,70).

En el primer escenario, el rango plausible apunta de forma consistente a un efecto positivo,

mientras que, en el segundo, la incertidumbre es mucho mayor y sería no solo más prudente, sino recomendable, replicar el estudio o complementar la evidencia antes de escalar una futura intervención.

Relevancia práctica, alta variabilidad y políticas para los ODS

Es necesario tener presente que ni el tamaño del efecto, ni mucho menos los intervalos de confianza, por sí solos, van a garantizar la toma de decisiones sensatas o correctas. Por ello, es necesario incorporar la noción de relevancia práctica, es decir, el juicio sobre si un efecto de cierta magnitud que cuenta con determinada incertidumbre es lo suficientemente importante para justificar, por ejemplo, una acción, una inversión o un cambio de política.

Y es que, en ámbitos como la salud pública o la educación, una mejora pequeña desde la perspectiva estadística puede ser altamente significativa en términos de vidas salvadas, reducción de desigualdades, de la exclusión o de la

equidad. Mientras que un efecto mediano podría resultar incluso irrelevante si los costos de implementación son demasiado altos.

Los contextos VUCA (por sus siglas en inglés: volatilidad, incertidumbre, complejidad y ambigüedad) de alta variabilidad, característicos de muchos problemas asociados a los ODS, añaden una capa adicional de complejidad. Aquí no es raro que estudios con muestras limitadas y que además presenten una gran heterogeneidad no alcancen $p < 0,05$, aun cuando los tamaños de efecto apunten a beneficios potencialmente relevantes.

En este sentido, es importante no descartar automáticamente resultados no significativos, sino situarlos en un continuo de evidencia donde la suma de esa combinación de tamaño de efecto, incertidumbre y plausibilidad teórica permita orientar la decisión sobre replicar, ajustar o escalar una determinada intervención, de acuerdo con Amrhein et al (2019) y Aguilar-Altamirano et al (2024).

En el marco del ODS 17, pero en especial, en lo atinente a la meta 17.19, se subraya la necesidad de fortalecer la capacidad estadística para desarrollar mediciones del avance en el desarrollo sostenible que trasciendan y complementen el PIB, permitiendo monitorear de una manera rigurosa la Agenda 2030 (Naciones Unidas, 2025).

Esto supone sistemas de información capaces de producir no solo indicadores o KPI puntuales, sino también estimaciones con bandas de incertidumbre transparentes y marcos de interpretación que integren la relevancia práctica.

Por ejemplo, un indicador de pobreza, de aprendizaje en una comunidad rural o de emisiones de gases de efecto invernadero no debería reportarse únicamente como un porcentaje, sino acompañado de su intervalo de confianza y de una reflexión sobre si las variaciones observadas son sustantivamente significativas y coherentes con las metas de la política de desarrollo.

Inferencia práctica y capacidades estadísticas para el desarrollo sostenible

En este punto, avanzar hacia una inferencia práctica basada en tamaños de efecto, intervalos de confianza y relevancia contextual implica también, aunque no lo parezca, un gran desafío formativo e institucional.

En términos de educación estadística, es necesario que investigadores, funcionarios públicos, empresas y profesionales de distintas disciplinas sean capaces de leer un intervalo de confianza, entendiendo la lógica de los tamaños de efecto y cuestionando las interpretaciones ingenuas que a menudo se hacen del p-valor.

Desde una perspectiva mucho más institucional, la meta 17.19 del Objetivo de Desarrollo Sostenible 17 demanda invertir en sistemas de información que puedan generar series temporales confiables, censos regulares, encuestas de elevada calidad, así como plataformas abiertas de datos que, apalancadas en la ciencia abierta, permitan escrutar

la incertidumbre presente detrás de cada indicador, según las Naciones Unidas (2025).

Estamos en presencia de un cambio que no es meramente técnico, se trata de uno que lleva inmerso un profundo sentido democrático y de justicia epistémica al hacer visible la incertidumbre reconociendo los límites propios del conocimiento y, al mismo tiempo, compartiendo con la ciudadanía y los múltiples actores sociales las condiciones reales bajo las cuales efectivamente se toman las decisiones.

Es que en lugar de presentar simples cifras como verdades absolutas, la combinación de las estimaciones, bandas de confianza y análisis de tamaño del efecto invita a repensar las políticas como apuestas informadas que están sujetas a revisión a la luz de nuevas evidencias.

Es necesario ir más allá del p-valor; la inferencia estadística orientada a la toma de decisiones en el marco del desarrollo sostenible exige de un enfoque centrado en la magnitud de los

efectos, de la explicitación de la incertidumbre, pero, ante todo, de la valoración de la relevancia práctica contextualizada.

Como se ha podido apreciar, el hallazgo central de este capítulo esta referido a que: es necesario interpretar tamaños de efecto junto con la incertidumbre y no solo la simple significancia, pues de este modo se producirán conclusiones más útiles para las decisiones reales, en especial, en escenarios marcados por una alta variabilidad y recursos limitados.

Este giro metodológico propuesto es coherente con los debates contemporáneos que se están gestando en materia de estadística aplicada, no en vano, consideran las recomendaciones de organismos como la ASA y centenares autores e investigadores estadísticos que critican la hegemonía del p-valor.

Sobre todo, es consistente con el espíritu implícito del ODS 17 y su meta 17.19, que reclaman la creación de indicadores de avance que sean cada

vez más robustos, transparentes y ante todo, contextualizados.

A partir de la adopción de una inferencia práctica, los actuales sistemas de investigación, los múltiples observatorios de políticas públicas y a su vez, las instituciones especializadas en estadísticas pueden ofrecer a los tomadores de decisiones algo mucho más valioso que un asterisco de un $p < 0,05$. Se trata entonces de ofrecer una comprensión rica de los efectos, de su incertidumbre y de su pertinencia para construir un desarrollo verdaderamente sostenible en beneficio de la humanidad.

III. DE LA ASOCIACIÓN A LA CAUSALIDAD EN INVESTIGACIÓN CUANTITATIVA

Car-Emyr Suescum Coelho[3]

La investigación cuantitativa con frecuencia utiliza la correlación como recurso para describir relaciones entre las variables. Esta práctica resulta valiosa, pese a que suele acompañarse de un desajuste entre el resultado de los cálculos y lo que se afirma, debido a que se atribuye una causalidad a patrones que solo reflejan covariación.

En este sentido, la credibilidad de un estudio de investigación se incrementa cuando se formula la pregunta con precisión, alineada con el método y el lenguaje de los resultados, tal como lo señalan Ito et al (2025). Esa alineación establece el alcance

[3] Doctor en Gerencia. Docente Titular de la Universidad Metropolitana, Venezuela y de la Universidad Central de Venezuela. csuescum@unimet.edu.ve ORCID: https://orcid.org/0000-0003-1104-7800

interpretativo del coeficiente y reduce la ambigüedad en la presentación de hallazgos.

En correspondencia el capítulo propone criterios explícitos de selección e interpretación de coeficientes de correlación, acompañados por reglas de redacción que eviten las extrapolaciones causales no sustentadas.

Se aborda el tema con la finalidad de fortalecer la calidad de los manuscritos empíricos, en los que persisten controles añadidos sin justificación y conclusiones que confunden asociación con efecto. Según Rohrer (2018) el razonamiento gráfico aporta una vía clara para exponer supuestos sobre confusión, selección y variables intermedias. De ese modo, el análisis correlacional queda situado en su función descriptiva y el lenguaje causal se reserva a contextos defendibles.

El texto se fundamenta en la revisión narrativa y analítica de literatura seleccionada, con lectura comparada de definiciones, supuestos y recomendaciones editoriales. La síntesis privilegia

criterios de decisión y consecuencias prácticas en la escritura académica, sin depender de demostraciones extensas.

La estructura se organiza en tres secciones de extensión similar, que avanzan desde la formulación del problema y del tipo de relación, hacia la elección del coeficiente y finalmente, hacia la interpretación y el reporte responsable. Esta secuencia se apoya en la idea de contrastar argumentos y evidencia desde enfoques distintos, en consonancia con lo propuesto por Hammerton & Munafò (2021).

Del problema científico a la asociación o causalidad

En toda investigación cuantitativa, la delimitación del problema científico condiciona la elección de los métodos y la interpretación de los resultados. El primer paso consiste en diferenciar si la intención es describir, predecir o explicar una relación entre variables.

De acuerdo con Ito et al (2025) la coherencia entre el tipo de pregunta y la estrategia analítica es la base para evitar inferencias erróneas. Cuando el propósito es únicamente descriptivo o predictivo, las correlaciones permiten identificar patrones o asociaciones, pero no justifican afirmaciones sobre mecanismos causales.

En la siguiente figura se realiza un breve recordatorio sobre las intenciones que persiguen los distintos tipos de investigación. La correcta definición de este aspecto resulta indispensable para seleccionar posteriormente el camino estadístico correcto para alcanzar los objetivos planteados.

Figura 3. Tipo de investigación según el propósito
Fuente: Elaboración propia, Suescum (2025).

La figura anterior muestra que el propósito del estudio delimita el alcance inferencial. Cuando se busca describir o predecir, la correlación orienta patrones; cuando se explica, exige supuestos explícitos claramente establecidos.

La confusión frecuente entre correlación y causalidad ha generado una tradición de interpretaciones inadecuadas que afectan la validez de muchos estudios observacionales. Foong & Vicendese (2025) destacan que una correlación significativa solo representa una covariación estadística entre variables, sin que esto implique dirección o influencia.

Para aproximarse a la causalidad, se requiere un razonamiento que incluya la temporalidad, la exclusión de causas alternativas y la coherencia teórica. No obstante, este enfoque no elimina la utilidad de las correlaciones, pero redefine su función como instrumentos descriptivos dentro de un proceso inferencial más amplio.

El análisis de correlaciones demanda reconocer la estructura de supuestos que sostiene cada interpretación. Rohrer (2018) explica que los modelos causales gráficos permiten representar las relaciones hipotéticas entre variables y visualizar las posibles fuentes de confusión.

Estos recursos facilitan la diferenciación entre un vínculo observado y si este realmente responde a un efecto real o a la acción de un factor común no controlado. Por consiguiente, el uso de diagramas causales favorece una reflexión más transparente sobre lo que los datos pueden respaldar o no pueden respaldar.

El razonamiento causal, también depende de la calidad de los datos y del diseño del estudio. Hammerton y Munafó (2021) argumentan que la triangulación de evidencia es indispensable para reforzar la credibilidad de los hallazgos, sobre todo cuando se emplean fuentes observacionales.

Esta estrategia combina resultados de distintos métodos o poblaciones para contrastar la

consistencia de las relaciones encontradas. De esta forma, el análisis de asociación se convierte en un punto de partida para examinar hipótesis causales bajo una lógica de convergencia empírica.

Por otra parte, las reflexiones metodológicas sobre la diferencia entre asociación y causalidad han evolucionado hacia modelos más integradores. Liuyi et al (2021) describen la forma en que la inferencia causal contemporánea se apoya en marcos formales, como los modelos de variables estructurales y los grafos dirigidos, que amplían las posibilidades de interpretación más allá del coeficiente de correlación.

La identificación de caminos causales potenciales requiere de una conceptualización precisa de los supuestos, de modo que la correlación no sea considerada como la evidencia final, sino como una medida preliminar dentro de un análisis más exigente.

Por consiguiente, la relación entre variables puede entenderse como un continuo entre la

descripción y la explicación. En este particular, Nogueira et al (2022) subrayan que el propósito no es eliminar el valor de las correlaciones, sino integrarlas en una secuencia analítica donde su papel se encuentre claramente delimitado.

De este modo, la claridad conceptual y la coherencia metodológica permiten evitar interpretaciones exageradas y promover conclusiones basadas en la lógica de los datos. Por su parte, Vallverdú (2017) complementa esta visión al advertir que la falta de precisión epistemológica puede convertir la correlación en un factor de confusión, especialmente cuando se mezcla con juicios causales sin fundamento empírico.

En este contexto, comprender la diferencia entre asociación y causalidad no implica una distinción puramente técnica, sino una decisión epistemológica que orienta la forma de investigar. Quispe et al (2020) afirman que los sesgos de confusión surgen cuando no se reconocen las variables que influyen simultáneamente en la causa

y el efecto aparente, distorsionando la interpretación de la relación.

De esto se desprende que el análisis de correlación requiere un control conceptual y estadístico que garantice coherencia entre la teoría y la evidencia.

Tal como lo proponen Runge et al (2019) el estudio de relaciones causales en conjuntos de datos complejos debe abordarse desde un enfoque que combine herramientas estadísticas y modelos no lineales. Este planteamiento, refuerza la necesidad de entender la correlación como parte de un sistema analítico más extenso, en el que los supuestos sobre independencia, temporalidad y direccionalidad sean explícitos.

Por tal razón, la precisión en la formulación del problema científico y la claridad en los objetivos de análisis constituyen el punto de partida para establecer una frontera metodológica sólida entre asociación y causalidad.

Elección del coeficiente de correlación: marco de decisión

Una vez delimitado el tipo de relación que se busca estudiar, la elección del coeficiente de correlación deja de ser un acto rutinario y se transforma en una decisión metodológica que debe quedar justificada. Según Ito et al (2025) el valor de un estudio aumenta cuando la pregunta, el método y el lenguaje de los hallazgos mantienen congruencia. Por esta razón, seleccionar un coeficiente adecuado requiere declarar si se pretende describir una asociación, comparar patrones o sustentar una interpretación explicativa bajo supuestos explícitos.

El punto de partida para esta selección es la naturaleza de las variables. Cuando son continuas y el interés se centra en una relación aproximadamente lineal, el coeficiente de Pearson suele ser el referente por su interpretación directa.

Si la relación esperada es monotónica o existen valores extremos que alteran la media y la varianza, los coeficientes basados en rangos, como Spearman

o Kendall, ofrecen una alternativa más estable para describir ordenamientos consistentes. En este sentido, la selección debe responder a la forma de dependencia que el problema plantea y no al hábito disciplinar.

La medición también determina los pasos subsiguientes. Cuando se trata de variables ordinales, la correlación basada en rangos se ajusta mejor a la información disponible porque evita suponer distancias iguales entre categorías.

En el caso de las variables dicotómicas, la asociación puede representarse mediante coeficientes diseñados para esa escala, siempre considerando el desequilibrio de frecuencias. Por su parte, Rohrer (2018) insiste en que las correlaciones no deben tratarse como pruebas de mecanismos, por lo que el nivel de medición y el proceso de generación de datos deben explicarse para evitar conclusiones que exceden la evidencia.

Es necesario destacar, que la forma de relación no siempre es lineal ni monotónica. En contextos con

umbrales, saturaciones o patrones curvilíneos, un coeficiente tradicional puede subestimar la dependencia o incluso sugerir ausencia de relación.

Runge et al (2019) demuestran que, en datos no lineales y series temporales extensas, la detección de asociaciones causales exige herramientas capaces de manejar estructuras complejas. En este sentido, antes de fijar un coeficiente, resulta pertinente examinar gráficamente la relación, evaluar su forma y reconocer si el fenómeno admite dependencias con rezagos o interacciones.

La estructura de dependencia entre observaciones también debe ser considerada. En mediciones repetidas o en datos longitudinales, calcular correlaciones como si cada registro fuera independiente distorsiona la magnitud real de la asociación.

Para Hammerton & Munafò (2021) la selección o pérdida de participantes en el tiempo puede alterar la relación estimada entre exposición y desenlace, con efectos que se confunden con cambios reales.

Por ello, cuando hay repetición o seguimiento, conviene usar enfoques que respeten la dependencia intra-unidad y reportar la forma en que se manejó el sesgo por selección.

Otro aspecto crítico, es la presencia de variables que influyen simultáneamente sobre las dos variables de interés. En palabras de Quispe et al (2020) un confusor puede producir una asociación falsa o modificar una asociación verdadera, lo que afecta cualquier coeficiente, incluso el más apropiado para la escala.

Sin embargo, el mensaje no es reemplazar la correlación por ajuste automático, sino reconocer que la decisión sobre correlación simple o condicionada debe apoyarse en teoría sustantiva y en una representación clara de las relaciones esperadas.

Además de esto, la inferencia causal contemporánea contribuye a ordenar estas decisiones sin convertir la correlación en un fin. Los autores Liuyi et al (2021) distinguen variables previas

al tratamiento de variables posteriores y muestran que, considerar ambas en el análisis conduce a sesgos que se reflejan en estimaciones engañosas.

En este orden de ideas, cuando la correlación se usa como un resumen preliminar, se debe indicar si las variables se interpretan como antecedentes, consecuencias o indicadores intermedios. Esta claridad evita correlaciones condicionadas que abren rutas ilusorias.

En aquellos escenarios donde se aspira a descubrir estructuras causales a partir de datos, las pruebas de independencia condicional adquieren un papel central y suelen apoyarse en medidas de asociación.

Nogueira et al (2022) describen algoritmos que orientan aristas a partir de relaciones de separación condicional, lo que exige rigor en cómo se estima dependencia y en qué condiciones se asume suficiencia causal. Por lo cual, cuando se menciona aprendizaje causal, es pertinente precisar si la

correlación se usa como medida descriptiva o como insumo para pruebas condicionadas.

Esto implica que la correcta selección de un coeficiente de correlación debe justificar la escala de las variables, la forma esperada de la relación, la estabilidad frente a valores extremos, la dependencia entre observaciones y el papel de posibles factores de confusión.

En este sentido, Foong & Vicendese (2025) afirman que cuando el objetivo declarado es causal, la selección del análisis debe alinearse con un razonamiento explícito sobre el efecto que se busca estimar. Bajo esa lógica, la correlación es valorada como medida de asociación, útil y limitada, dentro de una estrategia analítica coherente.

La siguiente tabla ofrece un criterio operativo para seleccionar la medida que describa la relación entre variables con coherencia metodológica. En este sentido, ordena decisiones que suelen tomarse por costumbre, vinculándolas con la escala de medición, la forma de dependencia esperada y el

diseño de los datos. Esta organización busca que el análisis sea consistente con el propósito del estudio y con el lenguaje usado al reportar resultados, tal como recomiendan Ito et al (2025).

Tabla 1
Medidas de asociación frecuentes

Condición del análisis	Medida de asociación sugerida	Criterio de elección y verificación mínima
Condición del análisis	Medida de asociación sugerida	Criterio de elección y verificación mínima
Dos variables continuas con patrón aproximadamente lineal	Pearson	Conviene corroborar linealidad con un diagrama de dispersión. Se reporta magnitud, dirección e intervalo de confianza.
Dos variables continuas con relación monotónica o presencia de valores atípicos	Spearman	El interés recae en el orden y no en la linealidad. Describir si hubo transformaciones por asimetría marcada.
Variables ordinales (Likert u ordenamientos)	Spearman o Kendall	Respeta el orden sin asumir distancias iguales entre categorías.
Una variable continua y una dicotómica	Punto-biserial	Verificar homogeneidad de varianzas y documentar la codificación de la variable dicotómica.
Dos variables dicotómicas	Phi	Revisar desbalance severo de frecuencias porque puede inducir magnitudes poco informativas.
Variables nominales con más de dos categorías	V de Cramér	Se sostiene en chi-cuadrado y resume asociación sin orden.

Dependencia no lineal con forma desconocida	Medida de dependencia no lineal	Se usa cuando Pearson o Spearman pueden ocultar relaciones curvilíneas.

Fuente: Elaboración propia, Suescum (2025).

A partir de la tabla anterior, conviene aclarar qué asociación y correlación no son equivalentes. La asociación alude a cualquier dependencia estadística entre variables, mientras que la correlación es una familia específica de coeficientes para cuantificar covariación bajo supuestos definidos.

Se empleó el término asociación en el encabezado porque el marco de decisión inicia antes de elegir un coeficiente y en ciertos casos la relación se resume mejor con medidas que no pertenecen a la correlación clásica. En este particular, Rohrer (2018) advierte que esa precisión reduce interpretaciones que exceden la evidencia.

La tabla anterior utiliza asociación como categoría general para evitar forzar una correlación en escenarios donde el tipo de variable o la estructura del dato sugieren otro resumen.

A su vez, cuando el texto se orienta hacia afirmaciones de efecto, la elección del análisis deja de depender solo del coeficiente y debe sostenerse en supuestos explícitos, constituyendo un requisito indispensable para cualquier lectura causal, en palabras de Foong & Vicendese (2025).

Interpretación y reporte responsable de correlaciones y causalidad

Es menester considerar que una correlación adquiere sentido científico cuando el manuscrito deja explícito qué tipo de pregunta que guía el análisis y qué inferencia se pretende sostener. Ito et al (2025) distinguen con precisión tres objetivos que suelen confundirse en artículos empíricos: describir patrones, predecir resultados o estimar efectos causales. Esa distinción exige coherencia entre pregunta, diseño, variables y lenguaje de resultados.

Cuando el texto utiliza verbos causales, la carga argumental deja de recaer en la magnitud del coeficiente y pasa a recaer en la interpretación contrafactual sobre lo que ocurriría sin la exposición,

bajo ciertas condiciones y con determinada población.

En este aspecto, Foong & Vicendese (2025) subrayan que no toda base observacional permite una interpretación causal defendible. Por consiguiente, el reporte debe indicar con claridad si el objetivo es asociación o efecto y cuáles supuestos vuelven plausible esa lectura.

Por otra parte, evitar el lenguaje causal no resuelve por sí solo el salto inferencial. La precaución retórica convive con un problema metodológico frecuente, descrito por Rohrer (2018) como el hecho de añadir controles sin justificación, con la creencia de que más control siempre mejora la inferencia, lo que a su vez puede degradar el razonamiento causal. En este sentido, el reporte debe justificar la selección de variables, explicitar supuestos y mostrar la estructura causal asumida, en lugar de confiar en advertencias genéricas.

Con relación a la interpretación, esta se distorsiona cuando un tercer factor induce una

asociación espuria entre exposición y desenlace. Quispe et al (2020) precisan que el sesgo de confusión se ha usado con varios significados.

Un confusor, en sentido amplio, es cualquier variable cuyo ajuste reduce sesgo en la estimación del efecto. Por tanto, la investigación debe declarar cuales variables fueron tratadas como confusoras, con qué criterio sustantivo y con qué estrategia de control.

Cuando un diseño único deja lugar a dudas, la lectura responsable exige contrastes que exploten sesgos distintos. Hammerton y Munafò (2021) describen la lógica de comparar poblaciones o grupos con estructuras de confusión diferentes.

Es decir, si el sesgo varía entre contextos, los resultados congruentes entre estos elevan la plausibilidad de un efecto causal. Mientras que si divergen, la hipótesis causal pierde fuerza. Esa idea obliga a reportar por qué los contextos difieren y qué sesgos se esperan en cada uno.

Es necesario destacar que el estándar de publicación en causalidad aplicada requiere trazabilidad computacional, no solo narración. Nogueira et al (2022) afirman que la inferencia causal no se reduce a correlaciones y que existe un amplio reservorio de métodos y herramientas.

Por consiguiente, un manuscrito debe detallar algoritmos, especificaciones, criterios de selección y decisiones de preprocesamiento, acompañado de códigos ejecutables y versiones, que faciliten la auditoría metodológica y la réplica.

En el caso de las series temporales, el reporte suele fallar al equiparar asociación con influencia. Runge et al (2019) enfatizan que detectar relaciones causales en datos temporales requiere más que solo medidas de correlación y debe incorporar estructura de rezagos, autocorrelación y posibles confusores latentes.

En otras palabras, publicar correlaciones significativas sin ese fundamento, induce lecturas causales injustificadas. La investigación debe

describir supuestos temporales, ventanas de rezago y pruebas de sensibilidad.

Incluso dentro del enfoque contrafactual, la validez depende de los supuestos y de su transparencia. Por su parte, Liuyi et al (2021) organizan métodos bajo el marco de resultados potenciales según el grado en que dependen de sus supuestos e incluyen recursos como datasets de referencia y código abierto.

Es recomendable que un reporte con pretensión causal, declare los supuestos adoptados, cómo se evalúan incumplimientos plausibles y qué evidencia empírica respalda la identificación del estimando.

Según Vallverdú (2017) la creatividad metodológica impulsa avances, pero durante su implementación genera confusión y controversia, sobre todo cuando las técnicas estadísticas se vuelven el centro del debate en campos aplicados.

Es decir, que la disciplina del reporte no es solo técnica; es epistemológica. El texto científico debe

sostener prudencia interpretativa, debe ser capaz de separar hallazgos descriptivos de pretensiones causales y dejar visibles los supuestos, sin convertir la novedad metodológica en argumento de verdad.

Como cierre de este capítulo, es necesario consolidar que los coeficientes de correlación aportan evidencia de asociación y su valor depende de la congruencia entre pregunta, datos y forma de reporte. Cuando se pretende inferir causalidad, esa pretensión exige explicitar el efecto de interés, la población a la que se refiere y los supuestos que permiten vincular datos observacionales con afirmaciones sobre cambios bajo intervención.

En ausencia de esa claridad, la correlación respalda descripciones y comparaciones, no afirmaciones sobre cambios atribuibles a una exposición.

Por esta razón, la calidad metodológica se fortalece al elegir el coeficiente adecuado, siendo transparentes con las decisiones analíticas y anticipando sesgos que inducen asociaciones

espurias. De esta manera, la relación entre asociación y causalidad queda delimitada y el reporte conserva precisión y credibilidad.

IV. RECOLECCIÓN DE DATOS PARA PRUEBAS DE HIPÓTESIS EN INVESTIGACIÓN

María Gabriela Gutiérrez Cadenas[4]

La prueba de hipótesis se ha presentado de manera clásica, como un conjunto de reglas de decisión basadas en estadísticos y probabilidades. Esta concepción resulta incompleta cuando se omite el hecho de que el contraste de hipótesis se fundamente en la forma en que se construyen los datos.

Cuando el levantamiento de información para la investigación incorpora definiciones ambiguas, mediciones inestables o registros incompletos, la decisión estadística corre el riesgo de describir artefactos del procedimiento en lugar de rasgos del fenómeno. Para Zhao & Velasco (2022) el

[4] Doctora en Química de Medicamentos. Profesor Agregado de la Universidad de Los Andes, Venezuela. mggutierrez239@gmail.com. https://orcid.org/0009-0004-2838-5448

planteamiento de hipótesis presupone la existencia de evidencia empírica pertinente y verificable.

El levantamiento de datos amerita la toma de decisiones previas que determinarán la calidad de la evidencia científica. En otras palabras, se requiere precisar variables, establecer reglas de medición, fijar protocolos de registro y asegurar la existencia de condiciones homogéneas de aplicación del método seleccionado.

En las investigaciones aplicadas, los problemas más comunes no se originan durante la ejecución de los cálculos, sino que resultan de la falta de trazabilidad del procedimiento, la variación introducida por los observadores y la pérdida de casos. La inferencia estadística se encuentra asociada con el control explícito de la variabilidad que acompaña al dato desde su origen, de acuerdo con lo planteado por Dagnino (2014).

Por otra parte, la coherencia metodológica se expresa en la relación lógica que vincula la pregunta, el instrumento, el muestreo y el análisis. Si la

variable se registra con categorías poco discriminantes, el efecto se disgrega. Además de esto, si la selección de unidades se realiza a conveniencia, la inferencia pierde el alcance, y si el protocolo se cambia durante el trabajo de campo, el dato deja de ser comparable.

En este sentido, Rodríguez et al (2012) plantean que la información que proporciona la prueba es de mayor calidad cuando se comprende como un proceso, con decisiones concatenadas que sostienen la interpretación.

De manera adicional, es importante destacar que el contraste de hipótesis funciona bajo riesgos de decisión. El nivel de significación no compensa sesgos de selección, ni la potencia aumenta por sí sola cuando la medición introduce algún ruido. La investigación requiere del establecimiento de reglas de calidad y de controles durante el levantamiento de información, de modo que el análisis sea el producto resultante de información consistente.

Para Sánchez et al (2024) la utilidad práctica del contraste de hipótesis disminuye cuando la prueba se usa como una rutina desconectada del procedimiento que genera los datos. En este orden de ideas, el presente capítulo tiene como finalidad examinar con criterios metodológicos y respaldo empírico, la forma potencial de perfeccionar la recolección de datos en pruebas de hipótesis mediante tres ejes articulados, acorde a lo propuesto por Leenen (2012).

El primer eje aborda la calidad del dato con énfasis en la validez, la confiabilidad y los sesgos. El segundo eje analiza el muestreo y el tamaño muestral, como decisiones que condicionan la precisión y el alcance inferencial. Finalmente, el tercer eje relaciona el levantamiento de información con errores tipo I y II y con potencia, de modo que la decisión estadística refleje verdaderamente el fenómeno, en lugar de las fallas del proceso.

Calidad del dato: validez, confiabilidad y control de sesgos

La calidad del dato se expresa en su correspondencia con el fenómeno que se pretende evaluar. La prueba de hipótesis supone que la variable observada representa una definición operacional coherente con la teoría y con el contexto del estudio.

Cuando esa correspondencia se debilita, el contraste puede conducir a decisiones que no guardan relación con el objetivo de investigación. Tal como destacan Zhao & Velasco (2022) la hipótesis no se basa en un cálculo aislado, sino en mediciones pertinentes y documentadas.

Por su parte, la validez se construye mediante evidencia que vincula indicadores con el concepto. Una definición operacional clara evita que la medición se desplace hacia dimensiones tangenciales o irrelevantes. Por ejemplo, en estudios que aplican cuestionarios, la validez de contenido se refuerza con la revisión por expertos y con pruebas piloto capaces de identificar oportunamente las ambigüedades y posibles sesgos de formulación.

En este contexto, Rodríguez et al (2012) afirman que la confusión conceptual se traduce en resultados sin interpretación consistente, aun cuando la prueba estadística se ejecute de forma correcta.

En relación con la confiabilidad, esta se refiere a la estabilidad del registro frente a repeticiones razonables del procedimiento. Una medición consistente, reduce el componente aleatorio del error y permite atribuir cierta variación al fenómeno evaluado.

La confiabilidad se fundamenta con la capacitación del investigador, el uso adecuado de manuales operativos, la calibración de instrumentos y la ejecución de ejercicios de concordancia cuando intervienen evaluadores. La inferencia depende de una variabilidad controlada; si la medición aporta ruido excesivo, el contraste pierde precisión, según lo expuesto por Dagnino (2014).

Por otra parte, la variación interobservador representa una fuente recurrente de deterioro de la evidencia. La inexistencia de criterios homogéneos

hace que la clasificación pueda cambiar según el evaluador, lo que a su vez introduce diferencias artificiales en los grupos comparados.

Por esta razón, resulta conveniente establecer reglas explícitas, ejemplos de casos límite y mecanismos de verificación durante el trabajo de campo. En este aspecto en particular, Mondal et al (2025) evidencian, desde el análisis de concordancia, que la estructura de la variación influye en el diseño y en las decisiones cuantitativas posteriores.

Es necesario destacar, que el control de sesgos involucra la anticipación de errores sistemáticos que pueden afectar las estimaciones. El sesgo de selección se produce cuando la inclusión depende de características asociadas con la variable de interés; mientras que el sesgo de información ocurre cuando el registro difiere según el grupo o la condición de exposición.

Con el objetivo de reducir estas distorsiones, el protocolo de investigación debe definir criterios de

inclusión consistentes y establecer procedimientos equivalentes de reclutamiento y medición. Según Sánchez et al (2024) el contraste de hipótesis pierde utilidad cuando el levantamiento introduce diferencias ajenas al fenómeno.

La estandarización del protocolo conlleva la toma de decisiones operativas verificables. Es necesario definir la unidad de observación, el momento de medición, las condiciones de aplicación, las reglas de codificación y formatos suficientemente explícitos para evitar omisiones.

Además, es recomendable un pilotaje previo para identificar posibles fallas de comprensión, tiempos de aplicación y puntos críticos del registro. La calidad inferencial se relaciona con el tratamiento integral del proceso investigativo, en el cual el dato se planifica junto con su interpretación, para evitar ajustes improvisados durante la recolección, acorde con lo señalado por Flores (2018).

Por otra parte, el manejo de datos faltantes debe planificarse antes del levantamiento. No basta con

reportar pérdidas; se necesita registrar causas, evaluar patrones de falta de respuesta y documentar las decisiones de reemplazo o de seguimiento. Si la pérdida de datos se asocia a variables clave, se genera un sesgo que afecta estimaciones y pruebas.

Según Dagnino (2014) la transparencia sobre variabilidad y pérdidas constituye parte del razonamiento inferencial, ya que define el significado del resultado obtenido. Por otra parte, la calidad de la investigación también depende del nivel de medición y de la escala utilizada. Las variables nominales, ordinales y continuas exigen instrumentos de medición distintos y generan riesgos de error diferentes.

La categorización excesiva puede reducir la sensibilidad y disminuir el efecto, mientras que una escala diseñada de manera incorrecta induce a reclasificaciones y variación artificial. Para Zhao & Velasco (2022) la formulación de hipótesis se asocia con la anticipación al tipo de evidencia que la

medición va a generar, lo que exige coherencia desde el diseño del instrumento de medición.

Con la finalidad de fortalecer la contrastabilidad, es recomendable ejecutar la auditoría del levantamiento de datos durante la propia recolección. Los controles de rango, la revisión de coherencias internas, la doble digitación y los análisis descriptivos tempranos permiten detectar errores de registro y patrones de omisión.

De acuerdo con Rodríguez et al (2012) estos controles operan como barreras que evitan que el análisis reciba un conjunto de datos con defectos estructurales y deben formar parte del procedimiento y no considerarlos una etapa opcional.

Otro aspecto de gran importancia en investigación es la documentación final, ya que constituye la base de la replicabilidad. Es necesario conservar evidencia de las versiones del instrumento, la capacitación aplicada, los criterios de codificación usados y las posibles modificaciones justificadas. Esta trazabilidad permite evaluar el

alcance de la prueba y la credibilidad de la decisión estadística.

Para Leenen (2012) gran parte de las críticas al contraste de hipótesis se relacionan con la confianza depositada en el resultado, adjudicadas a procedimientos de recolección poco transparentes. Por esta razón, la calidad del dato se entiende como condición de validez del contraste, no como complemento de este.

La muestra para pruebas de hipótesis

El muestreo define el alcance de la inferencia y la credibilidad de la comparación. Una prueba de hipótesis asume que la muestra representa con límites explícitos, a la población objetivo. Si su selección depende de conveniencia o de disponibilidad sesgada, el resultado puede describir una composición particular del grupo observado y no el fenómeno bajo estudio.

Según Zhao & Velasco (2022) el muestreo es un supuesto operativo del contraste de hipótesis y es la base de la generalización. En este sentido, la

elección del tipo de muestreo depende de la pregunta, del marco muestral y de la estructura de la población. El muestreo probabilístico ofrece condiciones más sólidas de inferencia, mientras que el no probabilístico exige declarar restricciones y evitar afirmaciones generalizadoras.

Cuando existen barreras de acceso, resulta adecuado describir la población objetivo, la población accesible y el procedimiento real de selección, con un debido registro de elegibilidad y rechazos. En este contexto, para Dagnino (2014) la aleatoriedad se asocia con el control de sesgos que afectan el razonamiento inferencial.

El muestreo estratificado resulta pertinente cuando se requiere la representación de subgrupos con diferencias esperadas. El estrato debe definirse en base a un criterio teórico y con viabilidad operativa, ya que un estratificado sin fundamento no aporta precisión y complica el diseño.

Además, es importante establecer cuotas, reglas de reemplazo y criterios de ponderación si las

fracciones muestrales difieren entre estas. Rodríguez et al (2012) señalan que la claridad del muestreo favorece la interpretación del contraste, dado que el investigador puede evaluar sesgos y límites de generalización.

Por otra parte, la determinación del tamaño muestral se sostiene en parámetros vinculados a la hipótesis. Es importante especificar la variable principal, la diferencia mínima de interés, la variabilidad esperada y el nivel de significación. Sin estos elementos, el tamaño muestral se convierte en una decisión arbitraria y el contraste de hipótesis pierde la dirección interpretativa.

La eficiencia inferencial se incrementa cuando el tamaño muestral se justifica con relación al efecto que se busca detectar y al error que se está dispuesto a asumir, acorde a lo afirmado por Flores (2018).

En cuanto a la variabilidad esperada, esta se estima mediante antecedentes, pruebas piloto o registros institucionales. Si se subestima la

dispersión, el tamaño calculado puede resultar insuficiente y se incrementa el riesgo de no detectar efectos reales. Por el contrario, si se sobreestima la dispersión, se consumen recursos sin que exista una mejora interpretativa.

Por lo antes expuesto, es importante documentar la fuente de la estimación y el razonamiento utilizado. En concordancia con los expuesto por Leenen (2012) una parte de los desaciertos del contraste proviene de decisiones mecánicas que ignoran el significado del efecto y su magnitud en el contexto del estudio.

En diseños de investigación con dependencia de observaciones, el tamaño muestral efectivo disminuye. Las muestras por conglomerados, las medidas repetidas y las estructuras jerárquicas ameritan el reconocimiento de la correlación intragrupo y el ajuste del cálculo. Si se ignora esa correlación, se sobreestima la información disponible y se subestima la varianza, lo que a su vez altera la potencia y ocasiona errores de decisión.

Para Mondal et al (2025) en el marco del coeficiente de correlación intraclase, el tamaño muestral requerido cambia según el modelo y el objetivo del contraste, lo cual evidencia la necesidad de alinear el diseño y la recolección de datos.

Por otra parte, la planificación debe contemplar las pérdidas. El tamaño inicial de la muestra debe incorporar un margen según tasas esperadas de abandono, junto con estrategias de seguimiento y de registro de motivos. Si la pérdida se concentra en un grupo o en un perfil específico, se genera un sesgo que afecta las estimaciones y la validez de la comparación.

Sánchez et al (2024) señalan que la prueba de hipótesis pierde utilidad cuando el diseño no controla las condiciones de campo que alteran la composición final de la muestra y debilitan la decisión.

La selección de unidades también involucra consistencia ética y metodológica. El reclutamiento debe evitar ejercer presiones o generar condiciones

asimétricas que distorsionen la participación. Desde el punto de vista inferencial, estas prácticas pueden resultar en una muestra con características sistemáticamente distintas de la población objetivo.

Según Rodríguez et al (2012) es necesario integrar la discusión de decisiones de campo en la formación estadística, de manera que se defina el alcance real del contraste más que en la fase de cálculo.

Además de esto, Leenen (2012) afirma que el contraste de hipótesis adquiere sentido cuando se integra con criterios sustantivos y con límites explícitos de interpretación. El tamaño muestral debe vincularse con relevancia sustantiva.

Un estudio puede detectar diferencias mínimas con muestras grandes, sin que ello justifique acciones o cambios. También puede fallar en detectar efectos de interés si el diseño queda subdimensionado. Por ello, se requiere definir una diferencia mínima pertinente antes del levantamiento y mantenerla en el argumento metodológico.

Errores tipo I y II

Los errores tipo I y tipo II representan riesgos inherentes a la decisión estadística y su magnitud resulta afectada por el diseño de recolección. El error tipo I corresponde al rechazo de una hipótesis nula verdadera, mientras que el tipo II corresponde a la no detección de un efecto real. Este marco conceptual se interpreta con propiedad cuando se reconoce que el control de estos riesgos se inicia en el levantamiento y no se limita al análisis, tal como afirman Zhao & Velasco (2022).

El nivel de significación fija una tasa de error tipo I bajo supuestos, aunque la práctica puede incrementarla por decisiones de campo y decisiones de análisis no planificadas. La prueba repetida de múltiples hipótesis sin control, la redefinición de variables tras revisar resultados o el cierre anticipado del levantamiento con base en resultados preliminares incrementan la obtención de hallazgos no confiables.

Leenen (2012) discute estas prácticas como fuentes frecuentes de interpretaciones erradas, dado que el contraste se usa como legitimación posterior y no como guía previa del estudio.

La potencia depende del tamaño muestral, la variabilidad y la magnitud del efecto. Un instrumento inestable incrementa la dispersión y reduce la potencia, aun con un tamaño moderado. Por esta razón, la calidad de la medición actúa como un factor determinante en la capacidad de detección.

Para Dagnino (2014) el razonamiento inferencial se debilita si la variabilidad observada proviene del procedimiento de medición y no del fenómeno, ya que se pierde el contraste efectivo.

La definición operacional condiciona la potencia. Si el indicador captura el fenómeno con poca sensibilidad, el efecto se atenúa y se incrementa la probabilidad de un no rechazo que no refleja la ausencia de efecto, sino la limitación de la medición. Por tal razón, es recomendable seleccionar escalas

y criterios con capacidad de registrar variación relevante.

La estructura de dependencia del dato puede alterar la potencia y los errores. En el caso de los diseños con medidas repetidas o con unidades agrupadas, es necesario reconocer la correlación intragrupo, ya que la información efectiva no equivale al conteo simple de observaciones.

Si se ignora esa estructura, se sobreestima la potencia y se minimiza el error tipo II de manera aparente. En este orden de ideas, Mondal et al (2025) muestran que la prueba sobre correlación intraclase exige alinear objetivo, modelo y tamaño, lo cual confirma que la recolección define la capacidad de prueba.

Además de esto, las pérdidas de datos reducen la potencia y pueden introducir sesgos. Un tamaño calculado asume un número final de observaciones, por lo que cualquier disminución reduce la capacidad de detección y puede alterar la composición de grupos si la pérdida es diferencial.

El control del error tipo I también exige disciplina en la definición de hipótesis. Se requiere declarar hipótesis primarias, establecer variables principales y fijar criterios de análisis antes del levantamiento. Cuando el estudio se guía por inspección sucesiva del dato y por cambios de criterio, el contraste pierde su función de control de error y se convierte en una herramienta confirmatoria.

Esta práctica explica parte del descrédito asociado a resultados significativos que no se replican, dado que el procedimiento no queda trazable, tal como advierte Leenen (2012).

Además de esto, la comunicación científica debe distinguir la ausencia de efecto de la falta de potencia. Un no rechazo puede reflejar un estudio subdimensionado, una medición con ruido o un efecto menor que el umbral de interés.

Por esta razón, Dagnino (2014) argumenta que resulta necesario reportar supuestos del tamaño, tasas de pérdida y razones metodológicas del nivel

de significación utilizado para valorar la credibilidad de la decisión estadística.

Por otra parte, la mejora de precisión en la medición puede ser más eficiente que un incremento marginal del tamaño. La capacitación de evaluadores, calibración y estandarización reducen el error aleatorio y aumentan la potencia sin ampliar el reclutamiento. Esta idea cobra especial peso cuando el principal componente de variabilidad proviene del procedimiento y no del fenómeno, según Zhao & Velasco (2022).

Es importante asumir que los riesgos de error se administran con reglas de levantamiento, con diseño muestral coherente y con disciplina analítica desde el protocolo. De acuerdo con Flores (2018) si estas condiciones no se cumplen, el contraste puede entregar decisiones con apariencia técnica, aunque con escaso valor empírico.

Resulta útil condensar la relación entre riesgos de decisión y controles de campo en un formato mínimo, para facilitar la auditoría metodológica

durante la recolección. La siguiente tabla sintetiza en términos operativos la forma en que ciertas decisiones del levantamiento inciden en el error tipo I, el error tipo II y la potencia, según el enfoque de Zhao & Velasco (2022).

Tabla 2
Riesgos de error y controles de recolección en pruebas de hipótesis

Riesgo	Efecto de la recolección	Control en campo
Error tipo I	Múltiples pruebas Decisiones tras ver resultados	Hipótesis única Regla de cierre Registro previo
Error tipo II	Muestra final baja Medición imprecisa	Sobre-reclutar Retener Estandarizar Medición
Potencia baja	Instrumento poco sensible Categorías reducidas	Piloto Escala adecuada Aplicación uniforme
Varianza inflada	Ruido del evaluador o equipo	Calibrar Capacitar Concordancia Doble medición
Dependencia ignorada	Agrupación o repetición no registrada	Identificadores Estructura Modelo previo
Pérdidas diferenciales	Abandono por perfil; sesgo de composición	Motivos Seguimiento Comparar perdidos/retenidos
Cambios de protocolo	Inconsistencia temporal o entre equipos	Control de versiones Bitácora Supervisión

Fuente: Elaboración propia, Gutiérrez (2025).

Como se observa en la tabla anterior, el error tipo I aumenta cuando el estudio permite comparaciones no planificadas o impide el levantamiento guiado por resultados parciales.

En cambio, el error tipo II se incrementa cuando el tamaño final cae por debajo del plan o cuando el instrumento añade variación por inconsistencias en la medición. En este sentido, la potencia no es una propiedad del software estadístico, sino una consecuencia del procedimiento de campo, como explica Dagnino (2014).

También se aprecia que la recolección debe registrar la estructura del dato para evitar supuestos errados de independencia. Si existen medidas repetidas o unidades agrupadas y no se documentan, se distorsiona la varianza y se altera la interpretación del contraste. Además, las pérdidas diferenciales pueden transformar la comparación en una selección de sobrevivientes, afectando la

credibilidad del resultado, como afirma Leenen (2012).

En síntesis, la tabla propuesta, funciona como una lista de verificación concentrada para alinear el levantamiento con la lógica de decisión. Cuando estos controles se ejecutan y se documentan, el contraste de hipótesis adquiere una base empírica más defendible y útil para decisiones en investigación aplicada, tal como proponen Sánchez et al (2024).

Finalmente, el presente capítulo ha situado la recolección de datos como condición metodológica para la prueba de hipótesis. La calidad del dato se sostiene en definiciones operacionales claras, consistencia de medición y control de sesgos, lo que permite fijar la relación entre teoría y registro empírico.

Cuando el levantamiento carece de trazabilidad, la decisión estadística describe fallas del procedimiento y no propiedades del fenómeno. En

otras palabras, la inferencia exige coherencia en todo el proceso y no solo en la etapa de cálculo.

También se ha argumentado que el muestreo y el tamaño muestral determinan el alcance de la generalización y la precisión del contraste. El diseño de selección, el tratamiento de pérdidas y el reconocimiento de la dependencia del dato condicionan la información efectiva del estudio.

El tamaño muestral depende del modelo y del objetivo del contraste, lo que exige justificar supuestos con transparencia y a su vez permite que el investigador valore sesgo, precisión y límites de interpretación.

En cuanto a los errores tipo I, tipo II y la potencia, se consideran riesgos que se administran desde el levantamiento. La disciplina en la definición de hipótesis, la estabilidad de medición y el control de pérdidas reducen distorsiones que el análisis posterior no puede corregir. De este modo, el capítulo destaca el hecho de que la credibilidad de la prueba se construye con decisiones de recolección,

documentadas y coherentes con el propósito del estudio.

V. ALGORITMOS AUMENTADOS: LA NUEVA FRONTERA DE LA INFERENCIA CUANTITATIVA TABULAR

Thania Jheny Torres Pernía[5]

En el paradigma contemporáneo de la ciencia de datos, la preponderancia de la predicción sobre datos tabulares se encuentra en conjunto con una demanda ascendente de inferencias más transferibles, calculadas y justificables, particularmente cuando la toma de decisiones demanda cuantificar incertidumbre y mantener interpretaciones causales. Este capítulo incorpora tres corrientes emergentes:

-La aparición de modelos fundacionales tabulares basados en el aprendizaje en contexto y redes ajustadas a prioridades (por ejemplo,

[5] Doctora en Educación. Universidad Autónoma de Chile, Chile. Thania.torres@uautonoma.cl. https://orcid.org/0000-0002-2947-7027

TabPFN), que transforman el esquema clásico de fit-predict hacia un régimen de inferencia amortizada;

-La consolidación de la inferencia causal aplicada potenciada por el aprendizaje flexible a través de arquitecturas explícitas (SEM/DAG/SCM) y el aprendizaje doble/debiased machine learning, que encapsulan el aprendizaje flexible en garantías inferenciales.

-La "agentización" del ciclo estadístico mediante agentes fundamentados en modelos de lenguaje, capaces de planificar, ejecutar y auditar flujos analíticos. Dispositivo híbrido y rastreable, una estructura en capas que integra la teoría (identificación), la Inteligencia Artificial (aprendizaje), la estadística (inferencia) y los agentes (orquestación) con el fin de incrementar la validez, la velocidad y la reproducibilidad en la investigación cuantitativa implementada.

Es así como, en la investigación del siglo XXI, el análisis cuantitativo enfrenta una paradoja interesante: se tiene más capacidad para hacer

predicciones y más datos disponibles que nunca, pero también hay una necesidad clara de justificar, revisar y compartir conclusiones en diferentes contextos.

Esto lleva a revisar y compartir conclusiones en diferentes escenarios, usando la inteligencia artificial, cuando el enfoque cambia de "predecir" a "inferir" (evaluando efectos, midiendo niveles de incertidumbre y tomando decisiones en situaciones de riesgo), la capacidad de los algoritmos, la cual debe estar sujeta a condiciones de validez y trazabilidad.

Este capítulo ofrece un resumen del estado actual de la investigación aplicada. Se enfoca en:

-La importancia de los datos en forma de tablas en la ciencia y las limitaciones de la metodología más utilizada, que son los árboles de decisión con apoyo de gradiente.

-Los modelos tabulares innovadores (especialmente, TabPFN) que cambian la forma de

hacer predicciones mediante el aprendizaje en contexto y la inferencia amortizada;

-La investigación causal contemporánea que usa ML/IA sin sacrificar las garantías de los resultados, mediante arquitecturas causales claras y el aprendizaje múltiple/debiased machine learning.

-Los agentes que utilizan modelos de lenguaje y cómo están cambiando el ciclo estadístico, lo que trae tanto nuevas oportunidades como riesgos para la reproducibilidad. De esta convergencia nace la idea del algoritmo aumentado como un dispositivo combinado y de varias capas para la inferencia cuantitativa.

Predicción en datos tabulares

Los datos en forma de tablas, organizados en filas y columnas, siguen siendo el formato más común para la evidencia cuantitativa en áreas como la biomedicina, la economía, la educación, el clima, la industria y las políticas públicas. En estos, la tarea principal es predecir un objetivo (ya sea mediante clasificación o regresión) usando un conjunto.

Esto se hace en condiciones que incluyen diferentes significados, tipos mixtos (continuos, ordinales, categóricos), datos faltantes, valores atípicos y un desequilibrio en las clases. Estos atributos, han ayudado considerablemente a la fuerza de los modelos que no son neuronales, especialmente los que se basan en árboles de datos.

Se puede estimar que, cerca de veinte años, las metodologías de aumento de declive se han vuelto el estándar para datos tabulares, con importantes implementaciones como XGBoost, LightGBM y CatBoost (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018).

Su éxito se debe a su buen desempeño en muestras pequeñas y medianas, su habilidad para trabajar con diferentes tamaños y distribuciones no uniformes, y su facilidad para adaptarse a situaciones prácticas. Sin embargo, el enfoque actual tiene limitaciones importantes para la inferencia cuantitativa aplicada.

1. Compartir conocimientos entre grupos a menudo es limitado: el modelo se adapta "por dataset" y el aprendizaje no se utiliza de manera constante.

2. Su rendimiento puede desmejorar cuando ocurre un cambio en la distribución y ante cambios temporales o contextuales. Esto es muy importante en estudios sociales, educativos y de salud.

En cuanto a su relación natural con las arquitecturas de redes neuronales y los mecanismos que se pueden diferenciar, es más complicada, lo cual limita su conexión con modelos generativos, representativos y multimodales.

Modelos básicos tabulares

La literatura reciente ha replanteado el problema tabular no solo como “elegir un modelo”, sino como aprender del propio algoritmo de aprendizaje. Este giro se expresa en el aprendizaje en contexto (in-context learning, ICL), modelos tipo transformer pueden resolver tareas a partir de ejemplos presentados en el contexto de entrada,

aproximando procedimientos algorítmicos sin entrenamiento adicional específico del dataset.

En esta línea se ubican las Prior-data Fitted Networks (PFNs), en lugar de entrenar un modelo para un conjunto específico, se entrena un transformer sobre millones de tareas/datasets sintéticos para que internalice un algoritmo reutilizable.

Operacionalmente, el modelo recibe simultáneamente filas etiquetadas (conocidas) y no etiquetadas (a predecir) y ejecuta “entrenar + predecir” en un único forward pass, desplazando el esquema clásico fit–predict hacia un régimen de inferencia amortizada. En términos epistemológicos, se pasa de ajustar parámetros por dataset a disponer de un procedimiento aprendido que actúa como infraestructura inferencial transferible (Nagler et al., 2023).

En este mismo sentido, se encuentra la corriente TabPFN, que se presenta como un modelo fundamentado para datos tabulares de pequeña a

mediana magnitud. En su desarrollo, el modelo se capacita únicamente en tareas sintéticas, provocadas por un prior generativo, de manera que aprende a "resolver" diversos problemas de tabulares a través del aprendizaje en contexto.

Hollmann et al (2025) informan que TabPFN logra un rendimiento superior en varias pruebas de clasificación y regresión con hasta 10,000 muestras, superando modelos de referencia sólidos basados en gradient boosting, incluso cuando estos aprovechan grandes presupuestos de ajuste.

Desde el punto de vista de la inferencia numérica, la contribución es doble. Por un lado, la amortización de inferencia baja mucho el costo de computación en el ciclo de modelado, validación y repetición, lo que es muy importante en áreas con limitaciones de tiempo y recursos.

En otro aspecto, al funcionar como un transformer generativo, el modelo tiene la capacidad de generar no solo estimaciones puntuales, sino también distribuciones predictivas que absorben la

incertidumbre y permiten la toma de decisiones probabilísticas.

El enfoque hacia la "distribución predictiva" reorienta el análisis cuantitativo hacia un futuro en el que calibración, riesgo y robustez se convierten en las salidas principales del pipeline (Hollmann et al., 2025). Conceptualmente, TabPFN puede ser entendido como una solución convencional a la problemática del pipeline.

Bajo un principio inducido por el proceso de generación sintética utilizado en el preentrenamiento, el modelo se aproxima a una distribución posterior predictiva de la variable objetivo, teniendo en cuenta los datos observados.

Sobre el diseño metodológico, esto quiere decir que el investigador ya no "acepta" un prior del software sin cuestionarlo. Ahora puede discutir y, al final, controlar el prior operativo al definir grupos de tareas sintéticas, complejidades y suposiciones que guían cómo funciona el sistema de inferencia.

Inferencia causal con el apoyo de ML/IA

La transición de la predicción a la inferencia causal implica un cambio de enfoque: la pregunta ya no es “qué se relaciona con qué”, sino “qué cambia (y en qué cantidad) si se interviene”. En este trabajo, Chernozhukov et al (2024) sugieren unir la inferencia causal con el aprendizaje automático (ML) y la inteligencia artificial (IA). Conectan los modelos estructurales clásicos (SEM) con sus versiones modernas en IA (DAG/SCM) y proponen el Double/Debiased Machine Learning (DML) como una herramienta para obtener inferencias válidas en situaciones con alta dimensionalidad y funciones complejas.

Una aportación fundamental es subrayar que la causalidad no se manifiesta de manera inherente al ajuste predictivo, requiere una arquitectura conceptual explícita que identifique mecanismos generativos, rutas de confusión, mediación y criterios de supervisión.

Esta aclaración permite diferenciar entre controles "buenos" y "malos", pensar sobre la

identificación y justificar qué estimación causal se está tratando de obtener y en qué condiciones, de acuerdo con Chernozhukov et al (2024).

El núcleo operativo

El enfoque se centra en DML, un grupo de métodos que permite obtener parámetros importantes de baja dimensión (como efectos promedio predictivos o causales), incluso cuando hay parámetros auxiliares de alta dimensión que se aprenden con herramientas de ML.

Entre los resultados obtenidos con herramientas de ML hay dos elementos clave: los scores ortogonales de Neyman, que hacen que el estimador sea localmente insensible a errores en la estimación de nuisance, y el cross-fitting, que utiliza una división de muestras para evitar sesgos por sobreajuste y mantener propiedades asintóticas que permiten intervalos de confianza y pruebas.

Formación causal actual un núcleo duro del razonamiento hipotético

Si la inferencia causal mejorada por ML/IA indica una nueva forma de innovar en métodos, el área también necesita textos que establezcan fundamentos y criterios de validez. En esa función se encuentra A First Course in Causal Inference de Ding (2024), que presenta el aprendizaje causal desde el enfoque de resultados potenciales, define estimaciones contrafactuales, indica mecanismos de asignación y obtiene estimaciones e inferencias que son coherentes con esos mecanismos.

Una reseña en Psychometrika señala que el libro organiza sus capítulos en seis secciones y dedica mucho tiempo a experimentos aleatorizados, explicando la inferencia de Fisher y de Neyman, la estratificación, la rerandomización, el ajuste por regresión y los diseños por pares, entre otros temas.

El texto también trata sobre estudios observacionales (como el propensity score y estimadores doblemente robustos), la evaluación de supuestos (sensibilidad y solapamiento), variables instrumentales y problemas después del tratamiento

(mediación y estratificación principal) y tiene un enfoque muy estadístico, según Rickles (2025).

Entonces, la idea de algoritmo aumentado tiene un núcleo duro que es importante: señala que el aumento algorítmico no solo consiste en automatizar cálculos, sino en mejorar la consistencia entre la definición del efecto, el diseño, la identificación, la estimación y la evidencia. En otras palabras, el aumento que es importante en la inferencia cuantitativa es el que hace que el razonamiento causal sea verificable y muestra claramente la debilidad de los supuestos.

El ciclo analítico agentificado

Al mismo tiempo, la práctica de la estadística está cambiando gracias a agentes que utilizan modelos de lenguaje (LLM-agents). Estos agentes entienden instrucciones en lenguaje natural, organizan subtareas, crean y ejecutan código, y presentan resultados de manera más fácil y rápida. Sun et al (2024) resumen este nuevo campo bajo el concepto de data agents, sugiriendo que pueden

cambiar las reglas del juego para facilitar el acceso y acelerar el análisis.

El Survey organiza el diseño de estos agentes en partes como planificación, razonamiento, reflexión, colaboración entre varios agentes, integración de conocimiento y creación de sistemas que usan herramientas (Python/R/SQL) en su entorno.

También se identifican desafíos importantes: reproducibilidad (variaciones por versiones y mensajes), confiabilidad (errores, alucinaciones, sobre confianza), evaluación (falta de estándares sólidos) y gobernanza (privacidad, seguimiento y control humano), de acuerdo con Sun et al (2024).

En la investigación cuantitativa aplicada, la verdadera ventaja no está solo en "hacer más rápido" el análisis, sino en transformar buenas prácticas en rutinas que se pueden seguir. Esto incluye la creación automática de reportes que se pueden repetir, auditorías de supuestos, controles

de calidad de datos, validación de modelos y documentación del proceso.

Sin embargo, el riesgo es el mismo, si el agente se enfoca en la rapidez de respuesta en lugar de la validez inferencial, puede crear historias que parecen científicas, pero tienen poco respaldo. Esto hace que sea esencial tener un diseño de guardrails metodológicos, sobre todo, cuando el análisis busca entender causas o tomar decisiones de política.

Resumen simplificado: el algoritmo aumentado

Con base en los tres ejes revisados, se sugiere definir el algoritmo aumentado como un dispositivo que combina cuatro capas complementarias de manera clara: (a) identificación (teoría causal y estimaciones), (b) aprendizaje (ML/IA para funciones complejas), (c) inferencia (garantías estadísticas) y (d) orquestación (agentes que ejecutan y supervisan el proceso). Su premisa es simple: ML/IA aprende lo complicado; la teoría causal hace que sea más fácil de entender; la estadística permite hacer

conclusiones y los agentes hacen que el proceso sea ejecutable, documentado y repetible.

Tabla 3
El algoritmo aumentado como un sistema

Capa	Propósito inferencial	Ejemplos de métodos /artefactos
Identificación	Declarar el estimado y los supuestos de validez (contrafactual, DAG/SCM, criterios de control).	Resultados potenciales (Ding, 2024); DAG/SCM (Chernozhukov et al., 2024).
Aprendizaje	Aprender funciones auxiliares o algoritmos reutilizables sin imponer formas funcionales rígidas.	Gradient boosting (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018); PFNs/TabPFN (Hollmann et al., 2025; Nagler et al., 2023).
Inferencia	Controlar sesgos de regularización/sobre ajuste y cuantificar incertidumbre con garantías.	DML: ortogonalización y cross-fitting (Chernozhukov et al., 2017; Chernozhukov et al., 2024).

Fuente: Elaboración propia, Torres (2025)

En esta arquitectura, la novedad no se reduce a incorporar “IA” como caja negra, sino a desplazar el centro del método hacia decisiones explícitas y

auditables. En tabulización, el investigador puede discutir el prior operativo y la calibración de incertidumbre; en causalidad, puede formalizar identificación y seleccionar estimadores con garantías (DML).

Así mismo, con agentes, puede institucionalizar la reproducibilidad mediante registros de ejecución y chequeos automatizados. El algoritmo aumentado funciona como mecanismo de democratización responsable, amplía capacidades sin diluir la rendición de cuentas metodológica.

Proyección

El estado del arte revisado evidencia un reordenamiento de la inferencia cuantitativa. En predicción tabular, la emergencia de modelos fundacionales como TabPFN sugiere que el aprendizaje puede desplazarse desde el “modelo por dataset” hacia algoritmos aprendidos y reutilizables, con inferencia amortizada y salidas probabilísticas.

En inferencia causal, la fusión con ML/IA se vuelve científicamente productiva cuando se integra a arquitecturas causales explícitas y a mecanismos como DML que preservan garantías inferenciales. En la práctica analítica, los agentes basados en LLM anuncian una agentificación del ciclo estadístico que puede acelerar, documentar y auditar flujos, siempre que existan guardrails metodológicos.

En este contexto, el algoritmo aumentado se propone como categoría operativa para institucionalizar el aumento: un diseño por capas que alinea identificación, aprendizaje, inferencia y orquestación.

La proyección inmediata es doble: (a) construir protocolos de investigación cuantitativa donde la incertidumbre sea salida central (no apéndice) y (b) desarrollar agentes que no solo escriban código, sino que razonen sobre supuestos, diagnósticos y criterios de validez, haciendo la inferencia más rápida, más transferible y, sobre todo, más justificable.

VI. EL NÚMERO DE HARSHAD-NIVEN Y SU APLICABILIDAD ESTADÍSTICA EN LA INVESTIGACIÓN

Yasmilys Graciela Briceño Pérez[6]

La matemática encierra un mundo aleatorio lleno de infinitas posibilidades donde se sitúa el macro y el microespacio caótico en la convergencia de series. Aquí destaca la teoría de números, a menudo considerada la rama más pura de las matemáticas, alberga conjuntos de enteros cuyas propiedades trascienden la mera curiosidad aritmética para ofrecer soluciones a problemas complejos en la ciencia de datos y la estadística aplicada.

Entre estas estructuras destacan los números de Harshad, definidos como aquellos enteros

[6] Doctora en Ciencias de la Educación. Posdoctora. Asesora legal de Somos Tesis Puerto Rico, Asesora académica de Multiacademia Internacional. Estados Unidos. Gerente de Asesoría Académica y Consultoría Gerencial BRIYAS, asesorayasmilys@gmail.com, https://orcid.org/0009-0006-5051-9304.

positivos que son divisibles por la suma de sus propios dígitos. “Un número natural se dice que es un número harshad o también de Niven si es divisible por la suma de sus dígitos. Por ejemplo, el número 18 es harshad porque es divisible por 1 + 8 = 9”, de acuerdo con Ibáñez (2020, p.1).

La historia de estos números es un testimonio de la democratización del conocimiento matemático; fueron conceptualizados originalmente en la década de 1950 por Dattatreya Ramchandra Kaprekar, un maestro de escuela indio conocido por su intuición recreativa pero profunda.

Kaprekar los denominó "Harshad", una palabra derivada del sánscrito que significa "dador de alegría”, reflejando la satisfacción estética que produce su simplicidad estructural, según Pajares, (2025, p.5).

En la perspectiva expuesta, no fue hasta 1977, durante una conferencia de matemáticas en Estados Unidos, que el matemático Iván Niven formalizó su estudio y presentó condiciones para su densidad

asintótica, razón por la cual en la literatura académica occidental a menudo se les refiere indistintamente como números de Niven, según Weisstein (s/f.).

Esta dualidad en su nomenclatura marca el punto de partida de un objeto de estudio que ha evolucionado desde los pasatiempos matemáticos hasta convertirse en un componente de interés para la generación de entropía en sistemas computacionales contemporáneos.

En este contexto, el objetivo principal de este capítulo es analizar las propiedades aritméticas y dinámicas de los números de Harshad-Niven para fundamentar su aplicabilidad estadística en la investigación científica. Se busca demostrar que la estructura no lineal de estos enteros permite superar las limitaciones de predictibilidad de los generadores lineales, proporcionando un marco metodológico para la obtención de secuencias pseudoaleatorias de alta calidad estadística.

Definición formal, características y propiedades aritméticas

Desde una perspectiva formal, un entero positivo n se considera un número de Harshad en una base numérica b si cumple la condición de congruencia $n \equiv 0 \ (m´od \ S_b(n)$, donde $S_b(n)$ representa la función suma de dígitos de n en dicha base.

Para ilustrar esto con un ejemplo concreto en el sistema decimal (base 10), consideremos el número 18. La suma de sus dígitos es 1 +8 = 9. Dado que 18 es divisible entre 9 (el resultado es 2), 18 es un número de Harshad. Por el contrario, el número 19 tiene una suma de dígitos de 1 + 9 = 10; al dividir 19 entre 10 obtenemos 1.9, un resultado no entero, lo que excluye al 19 de este conjunto.

Esta propiedad de divisibilidad no es estática; depende intrínsecamente de la base en la que se representa el número, lo que añade una capa de complejidad estadística. Por ejemplo, el número binario 111_2 (que equivale al 7 decimal) tiene una

suma de dígitos de 1 + 1 + 1 = 3; como 7 no es divisible entre 3, no es Harshad en base 2, aunque podría serlo en otras bases.

Un primer acercamiento es el desarrollo de investigaciones aplicadas en la ciencia para el análisis e interpretación de propiedades matemáticas, que avanzan a los hallazgos de una belleza geométrica, armónica y creativa. La expresión, por ejemplo, es que el año 2025 es el único año cuadrado perfecto que viviremos en esta generación, ya que aconteció en 1936 una y la próxima será para el año 2116, de acuerdo con Pajares (2025, p.6).

Estudios exhaustivos realizados por De Koninck y Doyon han profundizado en la distribución de estos números, demostrando que, aunque son infinitos, su densidad natural tiende a cero a medida que x tiende a infinito, comportándose de manera predecible bajo leyes logarítmicas, lo cual permite modelarlos estocásticamente, según De Koninck y Doyon (2019).

Una de las características más intrigantes para la investigación estadística es la capacidad de estos números para aparecer en secuencias consecutivas, rompiendo la intuición de que las propiedades de divisibilidad deberían estar aisladas. Cooper y Kennedy (1993) demostraron que existen secuencias de números de Harshad consecutivos de longitud arbitraria, aunque encontrar secuencias largas requiere números astronómicamente grandes.

Un ejemplo concreto y accesible de rachas de Harshad se encuentra en los enteros de una y dos cifras: la secuencia 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 es trivialmente Harshad (todo dígito divide a sí mismo y 10 divide a 1). Sin embargo, al observar números mayores, las secuencias se rompen y reforman.

Consideremos el bloque del 20 al 24: el 20 es Harshad (2 + 0 = 2, 20⁄2 = 10); el 21 es Harshad (2 + 1 = 3, 21⁄3 = 7); pero el 22 no lo es (2 + 2 = 4, 22⁄4 = 5,5), rompiendo la racha. Curiosamente, el 24 vuelve a serlo (2 + 4 = 6, 24⁄6 = 4).

Esta intermitencia estructural, donde la propiedad de ser Harshad, aparece y desaparece basándose en la fluctuación de la suma de dígitos; crea un patrón que, aunque determinista, emula características de procesos estocásticos, lo cual es vital para el modelado de sistemas que requieren variabilidad controlada, según Grundman (1994).

Estadística aplicada y la ruptura de la linealidad

La relevancia de los números de Harshad para la estadística aplicada y la investigación reside fundamentalmente en la naturaleza de la función suma de dígitos, S(n).

En la estadística computacional, uno de los mayores desafíos es evitar las correlaciones lineales no deseadas que surgen en los algoritmos deterministas, como los Generadores Congruenciales Lineales (LCG). La función S(n) actúa como un operador de perturbación no lineal.

A diferencia de las funciones lineales, donde una pequeña variación en la entrada produce una pequeña variación en la salida de forma

proporcional, la función suma de dígitos provoca cambios abruptos en la divisibilidad. Por ejemplo, al pasar del número 1999 al 2000, el valor numérico solo aumenta en 1, pero la suma de dígitos colapsa drásticamente de 28 (1 + 9 + 9 + 9) a 2 (2 + 0 + 0 + 0).

Este cambio radical altera las propiedades de congruencia del número, introduciendo una discontinuidad aritmética. Para un investigador que diseña simulaciones de Monte Carlo, incorporar esta dinámica de Harshad en los algoritmos de generación de números aleatorios ayuda a destruir las estructuras de retícula (lattice structures) que suelen sesgar los resultados en espacios multidimensionales, tal como advierten los estudios sobre generadores aleatorios robustos de L'Ecuyer (2012).

Un ejemplo de la existencia de varias estructuras interesantes, entre las que se cuenta una unión de intervalos discretos en el conjunto de los

números naturales, está dado por [26, 27] U [44, 46] U [59, 61] U [80, 82] U [112, 114].

Se observa que los puntos iniciales de estos intervalos (26, 44, 59, 80, 112) forman una sucesión no aritmética ni geométrica, pero sus diferencias finitas (18, 15, 21, 32) sugieren un posible polinomio de grado superior o una recurrencia no lineal.

Además, la secuencia incluye los números primos: 59, 61, 113. Números de Harshad (divisibles por la suma de sus dígitos): 27 (2+7=9, 27/9=3), 45, 60, 80, 81, 112, 114. Propiedades mixtas: compuestos, impares y pares, con una distribución que parece convexa en cada intervalo, según Pérez (2025).

La complejidad en sistemas deterministas

En el ámbito de la investigación cuantitativa contemporánea, la capacidad de modelar sistemas complejos es fundamental. Desde la simulación de fenómenos estocásticos mediante métodos de Monte Carlo hasta el diseño de protocolos criptográficos para la seguridad de la información, la

estadística aplicada depende intrínsecamente de fuentes de entropía o desorden controlado.

Sin embargo, las computadoras, siendo máquinas deterministas, no pueden generar azar verdadero; dependen de algoritmos matemáticos que simulan este comportamiento, conocidos como Generadores de Números Pseudoaleatorios (PRNG), de acuerdo con L'Ecuyer (2012).

Aquí es donde la teoría del caos ofrece un marco conceptual invaluable para la estadística. Un sistema caótico es aquel que, siendo determinista (regido por leyes fijas y sin variables aleatorias externas), presenta un comportamiento impredecible a largo plazo debido a su extrema sensibilidad a las condiciones iniciales, según Strogatz (2018).

Para el investigador en estadística aplicada, el caos determinista representa una oportunidad que permite construir modelos matemáticos que pasan las pruebas de validación estadística de aleatoriedad, pero que son totalmente reproducibles,

una característica esencial para la auditoría científica y la replicabilidad de experimentos.

La investigación actual busca alejarse de los modelos lineales simples, que a menudo fallan al capturar la complejidad de los datos reales, para adentrarse en dinámicas no lineales. En este contexto, las funciones aritméticas que introducen mezcla y confusión en los datos son candidatas ideales para desarrollar nuevos algoritmos de generación y modelado estocástico.

Números Harshad. Su aplicabilidad en la generación de entropía y criptografía

En el campo de la investigación avanzada, específicamente en criptografía y seguridad de datos, los números de Harshad ofrecen un mecanismo para diseñar funciones de dispersión (hashing) y generadores de flujo de claves.

Al utilizar la propiedad Harshad como una condición de filtro o como parte de una ecuación recursiva no lineal (por ejemplo, $x_{n+1} = x_n + S(x_n)^2$ (mo´d m), se pueden generar series temporales que

pasan pruebas estadísticas rigurosas de aleatoriedad, como la suite NIST SP 800-22.

La aplicabilidad concreta se observa en el diseño de S-Boxes (Cajas de Sustitución) dinámicas para algoritmos de cifrado de imágenes, donde la dependencia de la suma de dígitos asegura que la relación entre el texto plano y el cifrado sea altamente compleja y sensible a las condiciones iniciales.

Esto se alinea con los principios del caos determinista, un sistema simple basado en aritmética elemental puede producir una complejidad emergente suficiente para proteger la información contra ataques estadísticos diferenciales.

La investigación actual sugiere que hibridar mapas caóticos con aritmética de Harshad mejora la entropía de Shannon de los sistemas resultantes, acercándola al ideal teórico de 8 bits por byte, lo cual es un indicador de máxima incertidumbre e información.

Análisis estadístico de la dinámica caótica y el espacio de fases

Al proyectar los pares (x_n, x_{n+1}) en un diagrama de dispersión (Poincaré plot), se observa una ruptura de la simetría reticular típica de los generadores lineales. El diagrama no muestra patrones geométricos evidentes, sino una nube de puntos homogénea, indicativa de caos robusto, como se puede ver en la figura 4, seguidamente.

Aplicando la prueba de Kolmogorov-Smirnov (K-S) para una muestra de N = 10,000, el estadístico D se mantiene por debajo del valor crítico ($p > 0,05$). No existe evidencia estadística para rechazar la hipótesis nula de uniformidad. Esto implica que la distribución de los números generados es indistinguible de una distribución U (0, 1).

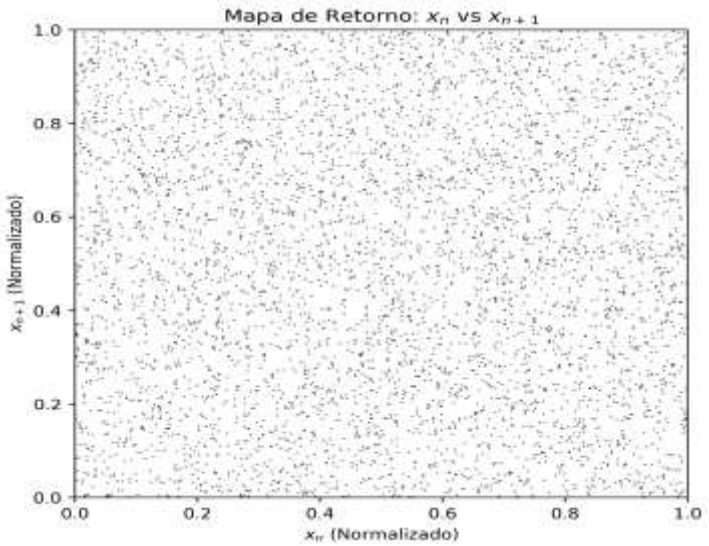

Figura 4. Diagrama de Poincaré para la secuencia MRH. Fuente: Marsaglia (1968).

Como se aprecia en la figura, la ausencia de patrones geométricos evidentes indica una ruptura exitosa de la simetría de retícula y una distribución uniforme en el hipercubo unitario. La introducción del término no lineal [S(xn)]2 provoca una dispersión homogénea en el espacio de fases unitario, indicativo de un comportamiento caótico robusto y buena independencia por pares.

Entre las aplicaciones en investigación estadística están: (a) generadores de números pseudoaleatorios (PRNG), viables para simulaciones Monte Carlo en ciencias sociales donde los generadores lineales fallan; (b) criptografía ligera, la irreversibilidad práctica de la función $S(x_n)$ permite enmascarar datos sensibles en bases de datos de investigación y (c) modelado de umbrales, útil para entender fenómenos naturales que dependen de la acumulación de recursos hasta un punto de quiebre (divisibilidad).

El exponente máximo de Lyapunov (MLE) mide la tasa promedio de divergencia de trayectorias cercanas. Para un sistema discreto unidimensional, se define como:

$$\lambda = (lím) \quad 1/N = \sum_{n=0}^{N-1} \ln[f'(xn)]$$

Dado que la función incluye la operación de suma de dígitos (que no es diferenciable suavemente), utilizamos el algoritmo de Wolf para series temporales.

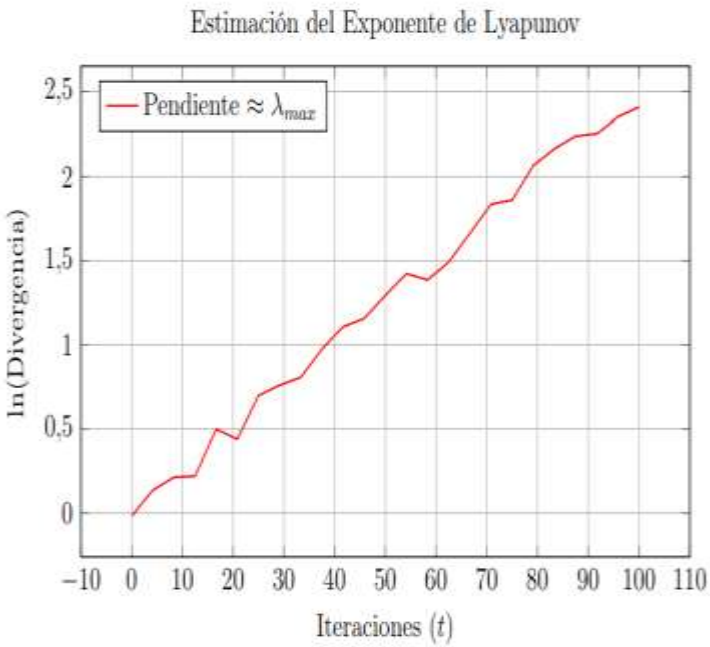

Figura 5. Divergencia logarítmica de trayectorias vecinas. Fuente: Elaboración propia, Briceño (2025)

Como se aprecia en la figura, la pendiente positiva de la regresión lineal indica λ max > 0, confirmando caos determinista. El análisis numérico arroja un valor de λ ≈ 2,45 (bits/iteración). Un valor positivo confirma que el sistema es caótico: la información sobre el estado inicial se pierde a una tasa exponencial, haciendo imposible la predicción a

largo plazo sin conocimiento exacto de los parámetros.

En conclusión, los números de Harshad-Niven representan mucho más que una curiosidad matemática sobre la divisibilidad. Su estructura interna, gobernada por la interacción no trivial entre el valor de un número y sus componentes digitales, ofrece un modelo determinista rico en complejidad.

Para la estadística y la investigación científica, estos números proporcionan una base teórica para comprender cómo surgen patrones complejos a partir de reglas simples y ofrecen herramientas prácticas para mejorar la calidad de las simulaciones computacionales y la seguridad de la información.

La capacidad de la función suma de dígitos para inyectar no linealidad y romper correlaciones convierte la dinámica de Harshad en un campo prometedor para el desarrollo de nuevas metodologías en la generación de números pseudoaleatorios y el análisis de sistemas dinámicos discretos.

REFERENCIAS

Aguilar-Altamirano, E. E., Bautista-Arroyo, E. Y & Rivera-Garamendi, F. G. (2024). *Confidence intervals. Multidisciplinary & Health Education Journal.* 6(1), 1002–1009. https://shre.ink/5hgy

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance*.* *Nature*. (7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9

Barrio, J. (2022). *Epistemología para maestros*. https://shre.ink/5hgz

Chen, T & Guestrin, C. (2016). *XGBoost: Ascalable tree boosting system*. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mininm (785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C & Newey, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *American Economic Review*,

107(5), 261–265. https://doi.org/10.1257/aer.p20171038

Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M & Syrgkanis, V. (2024). *Applied causal inference powered by ML and AI* (arXiv:2403.02467). arXiv. https://doi.org/10.48550/arXiv.2403.02467

Coe, R., & Merino-Soto, C. (2003). Magnitud del efecto: una guía para investigadores y usuarios. *Revista de Psicología*. 21(1), 145–177. https://doi.org/10.18800/psico.200301.006

Cooper, C. N & Kennedy, R. E. (1993). On consecutive Niven numbers. *Fibonacci Quarterly*, 31(2), 146-151. https://www.fq.math.ca/Scanned/31-2/cooper.pdf

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29. https://doi.org/10.1177/0956797613504966

Dagnino, J. (2014). Inferencia estadística: pruebas de hipótesis. *Revista Chilena de Anestesia*, 43, 125-128. https://shre.ink/qT1c

De Koninck, J. M & Doyon, N. (2019). The distribution of Niven numbers. *Journal of Number Theory*, 195, 12-28. https://doi.org/10.1016/j.jnt.2018.06.004

Díaz, C & García, Y. (2024). *El rol del investigador y la ética: la incansable lucha de vida.* Disponible en: https://doi.org/10.61286/e-rms.v2i.55

Ding, P. (2024). *A first course in causal inference*. Chapman & Hall/CRC. https://doi.org/10.1201/9781003484080

Escrig Sos, V. J., Miralles Tena, J. M & Martínez Ramos, D. (2007). Intervalos de confianza: por qué usarlos. *Cirugía Española*, 81(3), 121–125. https://doi.org/10.1016/S0009-739X(07)71281-2

Foong, R. & Vicendese, D. (2025). Considerations for Causal Inference Studies. *Respirology,* 30

(5), 369-454. https://doi.org/10.1111/resp.70018

Flores, P. (2018). Comparación de la eficiencia de las pruebas de hipótesis e intervalos de confianza en el proceso de inferencia. Estudio sobre medias. *Revista de Ciencias Naturales y Exactas*, 22(2), 65-85. https://doi.org/10.25100/rc.v22i2.7921

Grundman, H. G. (1994). Sequences of consecutive Niven numbers. *Fibonacci Quarterly*, 32(2), 174-175. https://shre.ink/5hgN

Hammerton, G., & Munafò, M. (2021). *Causal inference with observational data: the need for triangulation of evidence. Psychological Medicine*, 51, 563–578. https://doi.org/10.1017/S0033291720005127

Hollmann, N., Müller, S., Purucker, L et al. (2025). Accurate predictions on small data with a tabular foundation model. *Nature*, 637, 319–326. https://doi.org/10.1038/s41586-024-08328-6

Ibáñez, R. (2020). Los números que proporcionan alegría. *Cuaderno de Cultura Científica*. https://shre.ink/5hgV

Ito, C., Al-Hassany, L., Kurth, T & Glatz, T. (2025). Distinguishing description, prediction, and causal inference: a primer on improving congruence between research questions and methods. *Neurology*, 104(4), e210171. https://doi.org/10.1212/WNL.0000000000210171

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, 30.

L'Ecuyer, P. (2012). Random Number Generation. En J. E. Gentle, W. K. Härdle, & Mori, Y. (Eds.), *Handbook of Computational Statistics* (35-71). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21551-3_3

Leenen, I. (2012). La prueba de la hipótesis nula y sus alternativas: revisión de algunas críticas y su relevancia para las ciencias médicas. *Investigación en Educación Médica*, 1(4), 225-234. https://shre.ink/qT1n

Liuyi, Y., Zhixuan, C., Sheng, L., Yaliang, L., Jing G & Aidong, Z. (2021). A Survey on Causal Inference*. ACM Trans. Knowl. Discov*, 15 (5). https://doi.org/10.1145/3444944

López, F.; Casquete, N., Mejía, O y Barba, R. (2025). *Reconfiguración epistemológica de la estadística como soporte cognitivo en la era algorítmica: una revisión sistemática*. https://shre.ink/5hgO

Marsaglia, G. (1968). Los números aleatorios caen principalmente en los planos. *Actas de la Academia Nacional de Ciencias*, 61(1), 25-28.https://doi.org/10.1073/pnas.61.1.25

Mendivelso, F. O & Rodríguez, I. M. (2018). Definición e interpretación adecuada del p-valor e intervalos de confianza en investigación

clínica. *Revista Médica Sanitas*, 21(4), 193–196. https://shre.ink/5hgJ

Mondal, D., Cassese, A., Candel, M & Vanbelle, S. (2025). Sample size determination for hypothesis testing on the intraclass correlation coefficient in a two-way analysis of variance model. *British Journal of Mathematical and Statistical Psychology*, 00, 1-31. https://doi.org/10.1111/bmsp.70016

Naciones Unidas. (2025). *Metadata for SDG indicator 17.19.1: Dollar value of all resources made available to strengthen statistical capacity in developing countries*. United Nations Statistics Division. https://shre.ink/5hg0

Nagler, T., Janson, L., Bzdok, D., Athey, S & Bühlmann, P. (2023). *Statistical foundations of prior-data fitted networks.* arXiv. https://doi.org/10.48550/arXiv.2305.11022

Nogueira, A., Pugnana, A., Ruggieri, S., Pedreschi, D & Gama, J. (2022). *Methods and tools for*

causal discovery and causal inference. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 12(2), e1449. https://doi.org/10.1002/widm.1449

Pajares, P. D. (2025). 2025, un año muy matemático. *Unión - Revista Iberoamericana de Educación Matemática*, *21*(73). https://shre.ink/5hg6

Pérez, J.C. (2025). *Los números de Harshad*. Comunicación personal vía WhatsApp el 03 de noviembre de 2025.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In Advances in Neural Information Processing Systems, 31.

Quispe, A., Álvarez-Valdivia, M & Loli-Guevara, S. (2020). Metodologías Cuantitativas 2: Sesgo de confusión y cómo controlar un confusor. *Revista del Cuerpo Médico Hospital*

Nacional Almanzor Aguinaga Asenjo, 13(2), 205-212. https://shre.ink/5hg4

Ramos-Vera, C. (2020). *La inferencia bayesiana como replicación y cuantificación en la investigación clínica*. Disponible en: https://dx.doi.org/10.32641/rchped.vi91i6.3370

Rickles, J. H. (2025). A first course in causal inference. CRC Press [Book review]. Psychometrika. Cambridge University Press. https://shre.ink/5hgY

Rodríguez, M., Agnelli, H & Huerta, A. (2012). Pruebas de hipótesis estadísticas algunas consideraciones para la práctica docente. *Revista de Educación Matemática,* 1. https://shre.ink/qTUP

Rodríguez-Sabiote, C., Vázquez, L. M., López-López, J. A & Sánchez-Martín, M. (2025). Effect size as an alternative to statistical significance testing. A practical example with JASP. *Espiral. Cuadernos del Profesorado*, 18(38), 129–141. https://doi.org/10.25115/ecp.v18i38.10234

Rohrer, J. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science,*1(1), 27-42. https://journals.sagepub.com/doi/10.1177/2515245917745629

Runge, J. Nowack, P. Kretschmer, M. Flaxman, S & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5 (11), eaau4996. https://www.science.org/doi/10.1126/sciadv.aau4996

Sanchez, M., Marín, G., & Quintero, I. (2024). La importancia de la prueba de hipótesis. *Semilla Científica*, (5), 211-216. https://doi.org/10.37594/sc.v1i5.1381

Serrano-Díaz, G. E. (2022). La definición de probabilidad y su enseñanza a partir de sus significados: reflexiones. *Investigación y*

Postgrado, *37*(2),25–38. https://doi.org/10.56219/investigacinypostgrado.v37i2.1456

Strogatz, S. H. (2018). Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering. Westview Press. https://doi.org/10.1201/97804294925643

Sun, M., Han, R., Jiang, B., Qi, H., Sun, D., Yuan, Y & Huang, J. (2024). A survey on large language model-based agents for statistics and data science (arXiv:2412.14222). arXiv. https://doi.org/10.48550/arXiv.2412.14222

Toala, A.; Rodríguez, C.; Morales, J.; Quijano, A y Escobar, Y. (2024). *Inferencia estadística para enseñanza e investigación.* Disponible en: https://doi.org/10.37811/cli_w1102

Vallverdú, J. (2017). Creatividad epistemológica como factor de cambio y/o confusión: epidemiología y estadística. *Revista Iberoamericana de Argumentación*, 15, 92-98. https://doi.org/10.15366/ria2017.15.007

Wasserstein, R. L & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Weisstein, E. W. (s.f.). Harshad Number. En *MathWorld–A Wolfram Web Resource*. Recuperado de https://mathworld.wolfram.com/HarshadNumber.html

Zhai, C., Wibowo, S and Li, L.D. (2024). *Efectos de la excesiva dependencia de los sistemas de diálogo de IA en las capacidades cognitivas de los estudiantes: una revisión sistemática.* *https://doi.org/10.1186/s40561-024-00316-7*

Zhao, I & Velasco, M. (2022). Hypothesis Testing. *Revista MetroCiencia*, 30(1), 83-96. https://shre.ink/qTUY

Estadística transcompleja: inferencia, ética y algoritmos en la nueva ciencia

Leslibeth Joseliz Sucre Gonzalez

Carluys Suescum Coelho

Car-Emyr Suescum Coelho

Maria Gabriela Gutiérrez Cadenas

Thania Jheny Torres Pernia

Yasmilys Graciela Briceño Pérez

El texto es una cartografía que resignifica el enfoque mecanicista de la estadística tradicional y plantea una práctica renovada; deconstruye la ilusión de certeza absoluta hacia una nueva perspectiva que combina un enfoque riguroso de inferencia con vigilancia ética y un poder algorítmico ampliado. Además, ofrece herramientas esenciales para distinguir entre una simple asociación y una verdadera causalidad, haciendo énfasis en lo estadísticamente significativo, desde la epistemología de la incertidumbre hasta el uso de la inteligencia artificial y la teoría del caos.